



# High order discretization schemes for stochastic volatility models

Benjamin Jourdain, Mohamed Sbai

## ► To cite this version:

Benjamin Jourdain, Mohamed Sbai. High order discretization schemes for stochastic volatility models. The Journal of Computational Finance, 2013, 17 (2), pp.113-165. 10.21314/JCF.2013.262 . hal-00409861v4

**HAL Id: hal-00409861**

**<https://hal.science/hal-00409861v4>**

Submitted on 17 Oct 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# High order discretization schemes for stochastic volatility models

Benjamin Jourdain and Mohamed Sbai<sup>1</sup>

## Abstract

In typical stochastic volatility models, the process driving the volatility of the asset price evolves according to an autonomous one-dimensional stochastic differential equation. We assume that the coefficients of this equation are smooth. Using Itô's formula, we get rid, in the asset price dynamics, of the stochastic integral with respect to the Brownian motion driving this SDE. Taking advantage of this structure, we propose

- a scheme, based on the Milstein discretization of this SDE, which converges with order one to the asset price dynamics for an appropriate notion of convergence that we call weak trajectorial convergence,
- a scheme, based on the Ninomiya-Victoir discretization of this SDE, with order two of weak convergence to the asset price.

We also propose a specific scheme with improved convergence properties when the volatility of the asset price is driven by an Ornstein-Uhlenbeck process. We confirm the theoretical rates of convergence by numerical experiments and show that our schemes are well adapted to the multilevel Monte Carlo method introduced by Giles (*Multilevel Monte Carlo path simulation*. Operations Research, 56:607-617, 2008).

## Introduction

There exists an extensive literature on numerical integration schemes for stochastic differential equations. To start with, we mention, among many others, the work of Talay and Tubaro [29] who first established an expansion of the weak error of the Euler scheme for polynomially growing functions allowing for the use of Romberg extrapolation. Bally and Talay [4] extended this result to bounded measurable functions and Guyon [12] extended it to tempered stable distributions. More recently, many discretization schemes of higher weak convergence order have appeared in the literature. Among others, we cite the work of Kusuoka [18, 19], the Ninomiya and Victoir [25] scheme which we will use hereafter, the Ninomiya and Ninomiya [24] scheme and the scheme based on cubature on Wiener spaces of Lyons and Victoir [22].

Concerning strong approximation, the Milstein scheme has order one of strong convergence. Unfortunately, it involves the simulation of iterated Brownian integrals unless a restrictive commutativity condition is satisfied. Under ellipticity, Cruzeiro *et al.* [7] have recently proposed a discretization scheme which gets rid of these iterated integrals and has nice strong convergence properties. More precisely, for each number of time steps, there exists a Brownian motion different from the one giving the Brownian increments involved in the scheme such that the strong error between the scheme and the stochastic differential equation driven by this new Brownian motion is of order one. We call such a property weak trajectorial convergence of order one. Weak trajectorial error estimation is exactly what is needed to control the discretization bias for the computation of path dependent option prices.

Stochastic volatility models, which have now become a standard in the market, are an eloquent example of the use of stochastic differential equations in finance. In our study, we will consider the following specification

---

<sup>1</sup>Université Paris-Est, CERMICS, Projet MathFi ENPC-INRIA-UMLV. This research benefited from the support of the "Chaire Risques Financiers", Fondation du Risque and the ANR program BigMC.

Postal address : 6-8 av. Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 Marne-la-Vallée Cedex 2.

E-mails : [jourdain@cermics.enpc.fr](mailto:jourdain@cermics.enpc.fr) and [sbai@cermics.enpc.fr](mailto:sbai@cermics.enpc.fr)

which nests many current stochastic volatility models :

$$\begin{cases} dS_t &= rS_t dt + f(Y_t)S_t \left( \rho dW_t + \sqrt{1-\rho^2} dB_t \right); & S_0 = s_0 > 0 \\ dY_t &= b(Y_t)dt + \sigma(Y_t)dW_t; & Y_0 = y_0, \end{cases} \quad (1)$$

where  $(S_t)_{t \in [0, T]}$  is the asset price,  $r$  the instantaneous interest rate,  $(B_t)_{t \in [0, T]}$  and  $(W_t)_{t \in [0, T]}$  are independent standard one-dimensional Brownian motions,  $\rho \in [-1, 1]$  is the correlation between the Brownian motions respectively driving the asset price and the process  $(Y_t)_{t \in [0, T]}$  which solves a one-dimensional autonomous stochastic differential equation. The volatility process is  $(f(Y_t))_{t \in [0, T]}$  where the transformation function  $f$  is usually taken positive and strictly monotonic in order to ensure that the effective correlation between the stock price and the volatility keeps the same sign (the function  $\sigma$  usually takes nonnegative values). In the literature, the development of specific discretization schemes for stochastic volatility models has only received little attention. We mention nevertheless the work of Kahl and Jäckel [14] who discussed various numerical integration methods and proposed a simple scheme with order 1/2 of strong convergence like the standard Euler scheme but with a smaller multiplicative constant. Also the numerical integration of the CIR process and of the Heston model received a particular attention because of the inadequacy of the Euler scheme due to the fact that both  $f$  and  $\sigma$  are equal to the square root function (see for example Deelstra and Delbaen [8], Alfonsi [1], Kahl and Schurz [15], Andersen [3], Berkaoi *et al.* [5], Ninomiya and Victoir [25], Lord *et al.* [21], Alfonsi [2]). An exact simulation technique for the Heston model was also proposed by Broadie and Kaya [6].

In the present paper, we assume that the functions  $f$ ,  $\sigma$  and  $b$  are smooth which means that we do not deal with the Heston model where  $f(y) = \sqrt{y}$ ,  $b(y) = \kappa(\theta - y)$  and  $\sigma(y) = \nu\sqrt{y}$ . As an example of stochastic volatility models that fall within the scope of our study, let us mention

- Hull&White [13] model ( $\rho = 0$ ) and Wiggins [31] ( $\rho \neq 0$ )

$$\begin{cases} dS_t &= rS_t dt + \sqrt{Y_t}S_t \left( \rho dW_t + \sqrt{1-\rho^2} dB_t \right) \\ dY_t &= \mu Y_t dt + \zeta Y_t dW_t \end{cases}$$

which can be expressed as (1) with  $f(y) = \sqrt{y}$ ,  $b(y) = \mu y$  and  $\sigma(y) = \zeta y$ . Note that it can also be seen as (1) with  $f(y) = e^y$ ,  $b(y) = \frac{\mu}{2} - \frac{\zeta^2}{4}$  and  $\sigma(y) = \frac{\zeta}{2}$ .

- Scott's model [27] which generalizes the Hull&White model

$$\begin{cases} dS_t &= rS_t dt + \sigma_0 e^{Y_t} S_t \left( \rho dW_t + \sqrt{1-\rho^2} dB_t \right) \\ dY_t &= \kappa(\theta - Y_t)dt + \nu dW_t \end{cases} \quad (2)$$

$\Rightarrow f(y) = \sigma_0 e^y$ ,  $b(y) = \kappa(\theta - y)$  and  $\sigma(y) = \nu$ .

- Stein&Stein model [28]

$$\begin{cases} dS_t &= rS_t dt + Y_t S_t \left( \rho dW_t + \sqrt{1-\rho^2} dB_t \right) \\ dY_t &= \kappa(\theta - Y_t)dt + \nu dW_t \end{cases}$$

$\Rightarrow f(y) = y$ ,  $b(y) = \kappa(\theta - y)$  and  $\sigma(y) = \nu$ .

- Quadratic Gaussian model

$$\begin{cases} dS_t &= rS_t dt + Y_t^2 S_t \left( \rho dW_t + \sqrt{1-\rho^2} dB_t \right) \\ dY_t &= \kappa(\theta - Y_t)dt + \nu dW_t \end{cases}$$

$\Rightarrow f(y) = y^2$ ,  $b(y) = \kappa(\theta - y)$  and  $\sigma(y) = \nu$ .

Our aim is to take advantage of the structure of (1) to construct and analyse simple and robust ad hoc discretization schemes which have nice convergence properties. For a start, we make a logarithmic change of variables for the asset : the two-dimensional process  $(X_t := \log(S_t), Y_t)_{t \in [0, T]}$  solves the following SDE

$$\begin{cases} dX_t &= (r - \frac{1}{2}f^2(Y_t))dt + f(Y_t) \left( \rho dW_t + \sqrt{1 - \rho^2} dB_t \right); & X_0 = \log(s_0). \\ dY_t &= b(Y_t)dt + \sigma(Y_t)dW_t; & Y_0 = y_0. \end{cases} \quad (3)$$

Our main idea is to get rid in the first equality of the stochastic integral involving the common Brownian motion  $(W_t)_{t \in [0, T]}$ . In all what follows, we assume that

(A)  $f$  and  $\sigma$  are  $\mathcal{C}^1$  functions and  $\sigma > 0$ .

One can then define the primitive  $F(y) = \int_0^y \frac{f}{\sigma}(z)dz$  and apply Itô's formula to get

$$dF(Y_t) = \frac{f}{\sigma}(Y_t)dY_t + \frac{1}{2}(\sigma f' - f\sigma')(Y_t)dt.$$

Therefore  $(X_t, Y_t)_{t \in [0, T]}$  solves

$$\begin{cases} dX_t &= \rho dF(Y_t) + h(Y_t)dt + \sqrt{1 - \rho^2}f(Y_t)dB_t \\ dY_t &= b(Y_t)dt + \sigma(Y_t)dW_t \end{cases}, \quad (4)$$

where  $h : y \mapsto r - \frac{1}{2}f^2(y) - \rho(\frac{b}{\sigma}f + \frac{1}{2}(\sigma f' - f\sigma'))(y)$ . We discretize the autonomous SDE satisfied by  $Y$  using a scheme with high order of strong or weak convergence depending on whether one is interested in path-dependent or vanilla options. Then, in the dynamics of  $X$ , we only need to discretize the standard integral  $\int_0^T h(Y_s)ds$  and the stochastic integral  $\int_0^T f(Y_t)dB_t$  where  $(Y_t)_{t \in [0, T]}$  and  $(B_t)_{t \in [0, T]}$  are independent.

We recall that weak convergence is the right notion to analyse the discretization bias for plain vanilla options whereas weak trajectorial convergence permits to deal with path-dependent options. The first section of the paper is devoted to path-dependent options. Combining the Milstein discretization of the one-dimensional SDE satisfied by  $(Y_t)_{t \in [0, T]}$  with an appropriate discretization of the integral  $\int_0^T f(Y_t)dB_t$  based on the independence of  $(Y_t)_{t \in [0, T]}$  and  $(B_t)_{t \in [0, T]}$ , we obtain a scheme with order one of weak trajectorial convergence under several assumptions, the most restrictive one being that  $f^2$  is bounded away from 0. When  $(Y_t)_{t \in [0, T]}$  follows an Ornstein-Uhlenbeck process, which is the case for all the models cited above, the order one is preserved when replacing the Milstein discretization with exact simulation. Unfortunately, the assumption  $f^2$  bounded away from 0 is not satisfied by the stochastic volatility models cited above. For Scott and Hull & White models, where  $f$  is positive, we manage to prove that the order one is preserved. For the quadratic Gaussian model, taking advantage of the flatness of  $f(y) = y^2$  around the origin where this function vanishes, we are able to prove that the order of convergence is  $1 - \varepsilon$  for any  $\varepsilon > 0$ . Finally, in the Stein & Stein model, the fact that the derivative of  $f(y) = y$  does not vanish where this function is zero weakens the order of convergence : the order  $\frac{3}{4} - \varepsilon$  in  $L^2$  obtained by our theoretical analysis is confirmed by our numerical experiments.

In the second section, using the Ninomiya-Victoir discretization of the SDE satisfied by  $(Y_t)_{t \in [0, T]}$ , we construct a scheme with order two of weak convergence. Since the SDE satisfied by  $Y$  is one-dimensional, the Ninomiya-Victoir scheme only involves two one-dimensional ODEs whose solutions are available in closed form. The last section is devoted to numerical experiments which confirm the theoretical rates of convergence. We also compare the time needed by the different schemes to achieve a given precision in the multilevel Monte Carlo computation of a plain vanilla Call option and a lookback option. The multilevel Monte Carlo method proposed recently by Giles [9] automatically balances the bias and the statistical error and optimally takes advantage of both the weak and the strong convergence properties of the schemes to accelerate the computation. Somehow surprisingly, the strong convergence order has a dominating effect on its efficiency. We are able to exhibit an explicit coupling with order one of convergence between our weak trajectorial schemes with  $N$  and  $2N$  steps (see Remark 5). With this coupling, the multilevel Monte Carlo estimator

behaves as if the scheme had order one of strong convergence and the computation time needed to achieve the root-mean-square error  $\epsilon > 0$  is  $\mathcal{O}(\epsilon^{-2})$  (see [9]). For high levels of precision our schemes turn out to be more efficient than the Euler and the Kahl-Jäckel schemes for both the vanilla Call and the lookback option. The reason is that their better convergence properties compensate for the increase of computation effort at each step.

## Notations

We will consider, for a number of time steps  $N \geq 1$ , the uniform subdivision  $\Pi_N = \{0 = t_0 < t_1 < \dots < t_N = T\}$  of  $[0, T]$  with the discretization step  $\delta_N = \frac{T}{N}$ .

We denote by  $\underline{\psi}$  the greatest lower bound of the function  $\psi : y \mapsto f^2(y)$  and by  $\overline{\psi}$  its lowest upper bound. We also introduce the following notation :

$$\widehat{\psi}(y) = \begin{cases} \frac{3}{2}f^2(y) & \text{if } \overline{\psi} = \infty \\ \overline{\psi} & \text{otherwise.} \end{cases}$$

## 1 An efficient scheme for path dependent options pricing

Building a first order strong convergence scheme for a two dimensional SDE is not an obvious task. Even the ad hoc schemes provided by Kahl and Jäckel [14] exhibit a strong convergence of order  $\frac{1}{2}$ .

Actually, the natural candidate for this purpose is the Milstein scheme. Unfortunately, the commutativity condition which permits to implement it amounts to  $\sigma f' = 0$  in our setting. This condition is typically true when either  $f$  is constant or  $\sigma = 0$ . Both cases are of no practical interest since they lead to a deterministic volatility.

However, since the inherent Brownian motion is not essential for applications in finance, the usual strong convergence criterion is not crucial for estimating the error of a scheme in pricing a path dependent option. What is more relevant is the approximation in law of the whole trajectory of the process considered for instance by Cruzeiro *et al.* [7]. Using an ingenious rotation of the Brownian motion, these authors have constructed a discretization scheme allowing for a weak convergence on the whole trajectory of order one which avoids the simulation of the iterated stochastic integrals.

For the SDE (3), the discretization scheme of Cruzeiro, Malliavin and Thalmaier writes as

$$\begin{aligned} X_{t_{k+1}}^{CMT} &= X_{t_k}^{CMT} + \left( r - \frac{f^2(Y_{t_k}^{CMT})}{2} \right) \delta_N + \rho f(Y_{t_k}^{CMT}) \Delta W_{k+1} + \frac{\rho}{2} \sigma f'(Y_{t_k}^{CMT}) \Delta W_{k+1}^2 \\ &\quad + \sqrt{1 - \rho^2} \sigma f'(Y_{t_k}^{CMT}) \Delta W_{k+1} \Delta B_{k+1} + \sqrt{1 - \rho^2} f(Y_{t_k}^{CMT}) \Delta B_{k+1} - \frac{\rho}{2} \sigma f'(Y_{t_k}^{CMT}) \Delta B_{k+1}^2 \\ Y_{t_{k+1}}^{CMT} &= Y_{t_k}^{CMT} + \left( b(Y_{t_k}^{CMT}) + \frac{1}{2} \left( \frac{\sigma^2 f'}{f} - \sigma \sigma' \right) (Y_{t_k}^{CMT}) \right) \delta_N + \sigma(Y_{t_k}^{CMT}) \Delta W_{k+1} \\ &\quad + \frac{1}{2} \sigma \sigma' (Y_{t_k}^{CMT}) \Delta W_{k+1}^2 - \frac{\sigma^2 f'}{2f} \Delta B_{k+1}^2 \end{aligned} \tag{5}$$

where  $\Delta W_{k+1} = W_{t_{k+1}} - W_{t_k}$  and  $\Delta B_{k+1} = B_{t_{k+1}} - B_{t_k}$  correspond to the Brownian increments.

We set out to construct a much simpler scheme having the same order of weak trajectorial convergence by taking advantage of the particular structure of the SDE defining stochastic volatility models. We first begin with the general case of any process  $(Y_t)_{t \in [0, T]}$  driving the volatility and then consider the case of an Ornstein-Uhlenbeck process where we obtain more precise results.

### 1.1 General case

A discretization scheme will naturally involve the Brownian increments. Thanks to the independence between  $(Y_t)_{t \in [0, T]}$  and  $(B_t)_{t \in [0, T]}$ , we can construct a vector  $(\tilde{X}_{t_0}, \dots, \tilde{X}_{t_N})$  using only  $(\Delta B_1, \dots, \Delta B_N)$  and  $(Y_t)_{t \in [0, T]}$ , which has exactly the same law as  $(X_{t_0}, \dots, X_{t_N})$  :

**Lemma 1** —  $\forall 0 \leq l < N$ , let  $v_l = \frac{1}{\delta_N} \int_{t_l}^{t_{l+1}} \psi(Y_s) ds$ . The vector  $(\tilde{X}_{t_0}, \dots, \tilde{X}_{t_N})$  defined by

$$\begin{aligned} \tilde{X}_{t_0} &= X_{t_0} \\ \forall 1 \leq k \leq N, \tilde{X}_{t_k} &= \tilde{X}_{t_0} + \rho(F(Y_{t_k}) - F(Y_{t_0})) + \int_0^{t_k} h(Y_s) ds + \sqrt{1 - \rho^2} \sum_{l=0}^{k-1} \sqrt{v_l} \Delta B_{l+1} \end{aligned}$$

has the same law as  $(X_{t_0}, \dots, X_{t_N})$ .

*Proof :* The proof is elementary. Conditionally on  $Y$ , the two vectors are Gaussian vectors with the same mean and covariance matrix.  $\square$

In order to approximate  $(\tilde{X}_{t_k})_{0 \leq k \leq N}$ , one needs to discretize  $v_k$  for  $k \in \{0, \dots, N-1\}$ . If  $(v_k^N)_{0 \leq k \leq N-1}$  is an approximation of  $(v_k)_{0 \leq k \leq N-1}$ , then by Doob's inequality

$$\begin{aligned} \mathbb{E} \left[ \sup_{0 \leq k \leq N-1} \left( \sum_{l=0}^k \left( \sqrt{v_l} - \sqrt{v_l^N} \right) \Delta B_{l+1} \right)^2 \right] &\leq 4\delta_N \sum_{k=0}^{N-1} \mathbb{E} \left[ \left( \sqrt{v_l} - \sqrt{v_l^N} \right)^2 \right] \\ &\leq \frac{1}{\underline{\psi}} \delta_N \sum_{k=0}^{N-1} \mathbb{E} \left[ (v_l - v_l^N)^2 \right] \end{aligned}$$

as soon as  $\underline{\psi} = \inf_x \psi(x)$  is assumed to be positive and,  $\forall 0 \leq k \leq N-1$ ,  $v_k^N$  is greater than  $\underline{\psi}$ . Consequently, to obtain a scheme with order one of strong convergence for  $(\tilde{X}_{t_k})_{0 \leq k \leq N}$ , one needs that  $\forall 0 \leq k \leq N-1$ ,  $\mathbb{E} \left[ (v_k - v_k^N)^2 \right] = \mathcal{O} \left( \frac{1}{N^2} \right)$ . According to the treatment of the term  $\tilde{T}_2^j$  defined by (9) in the proof of the Theorem 2 below, one has  $\forall 0 \leq k \leq N-1$ ,

$$\mathbb{E} \left[ \left( v_k - \left( \psi(Y_{t_k}) + \frac{\sigma \psi'(Y_{t_k})}{\delta_N} \int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds \right) \right)^2 \right] = \mathcal{O} \left( \frac{1}{N^2} \right). \quad (6)$$

This equality still holds true when replacing  $Y$  by a scheme with order one of strong convergence in the term with sign minus of the left hand side. Better still,  $\left( F(Y_{t_k}) + \int_0^{t_k} h(Y_s) ds \right)_{0 \leq k \leq N}$  is approximated with strong order one when replacing  $Y$  by such a scheme and using a rectangular discretization for the integral in time.

For all these reasons, we choose the Milstein scheme for  $Y$  :

$$\forall 0 \leq k \leq N-1, \tilde{Y}_{t_{k+1}}^N = \tilde{Y}_{t_k}^N + b(\tilde{Y}_{t_k}^N) \delta_N + \sigma(\tilde{Y}_{t_k}^N) \Delta W_{k+1} + \frac{1}{2} \sigma \sigma'(\tilde{Y}_{t_k}^N) (\Delta W_{k+1}^2 - \delta_N); \quad \tilde{Y}_{t_0}^N = y_0.$$

and we write our scheme as follows

**WeakTraj-1 scheme**

$$\begin{aligned} \tilde{X}_{t_{k+1}}^N &= \tilde{X}_{t_k}^N + \rho \left( F(\tilde{Y}_{t_{k+1}}^N) - F(\tilde{Y}_{t_k}^N) \right) + \delta_N h(\tilde{Y}_{t_k}^N) \\ &\quad + \sqrt{1 - \rho^2} \sqrt{\left( \psi(\tilde{Y}_{t_k}^N) + \frac{\sigma \psi'(\tilde{Y}_{t_k}^N)}{\delta_N} \int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds \right) \vee \underline{\psi}} \Delta B_{k+1}. \end{aligned} \quad (7)$$

Note that in order to implement this scheme, one needs to simulate both the Brownian increment  $\Delta W_{k+1}$  and the random variable  $\int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds$ . This is straightforward as one can easily check that

$$\left( \begin{array}{c} \Delta W_{k+1} \\ \int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds \end{array} \right) \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \delta_N & \delta_N^2/2 \\ \delta_N^2/2 & \delta_N^3/3 \end{pmatrix} \right).$$

We can now state our first main result :

**Theorem 2** — *Under the following assumptions*

(H1)  *$f$  and  $\sigma$  are  $\mathcal{C}^3$  functions,  $\frac{f}{\sigma}$  and  $ff'$  are bounded*

(H2)  *$\underline{\psi} > 0$*

(H3) *there exists a constant  $K_1$  such that,  $\forall (x, y) \in \mathbb{R}^2$ ,*

$$\begin{aligned} \left| (bh' + \frac{\sigma^2}{2}h'')(y) \right| &\leq K_1(1 + |y|) \\ |\sigma h'(y)| &\leq K_1(1 + |y|) \\ |h(y) - h(x)| &\leq K_1|y - x| \end{aligned}$$

(H4) *there exists a constant  $K_2$  such that,  $\forall (x, y) \in \mathbb{R}^2$ ,*

$$\begin{aligned} \left| (b\psi' + \frac{\sigma^2}{2}\psi'')(y) \right| &\leq K_2(1 + |y|) \\ |\sigma\psi'(y) - \sigma\psi'(x)| &\leq K_2|y - x| \end{aligned}$$

(H5)  *$b$  and  $\sigma$  are  $\mathcal{C}^2$  functions with bounded first and second derivatives*

(H6) *there exists a positive constant  $K$  such that  $\forall (x, y) \in \mathbb{R}^2$*

$$|\sigma\sigma'(x) - \sigma\sigma'(y)| \leq K|x - y|$$

*the WeakTraj\_1 scheme has order one of weak trajectorial convergence. More precisely, for each  $p \geq 1$ , there exists a constant  $C$  independent of the number of time steps  $N$  such that*

$$\mathbb{E} \left[ \max_{0 \leq k \leq N} \left\| \left( \tilde{X}_{t_k}, Y_{t_k} \right) - \left( \tilde{X}_{t_k}^N, \tilde{Y}_{t_k}^N \right) \right\|^{2p} \right] \leq \frac{C}{N^{2p}}.$$

The proof of the theorem relies on the order one of strong convergence of the Milstein scheme (see Milstein [23] for the particular case  $p = 1$ ) :

**Lemma 3** — *Under the assumptions (H5) and (H6), one has that,  $\forall p \geq 1$ , there exists a positive constant  $C_p$  independent of  $N$  such that*

$$\mathbb{E} \left( \max_{0 \leq k \leq N} \left| Y_{t_k} - \tilde{Y}_{t_k}^N \right|^{2p} \right) \leq C_p \delta_N^{2p}.$$

The proof for general  $p$  is postponed to the appendix.

**Remark 4** — Before giving the proof of the theorem, we make a few comments on its assumptions.  $(\mathcal{H}1)$  implies that  $h$  and  $\psi$  are  $\mathcal{C}^2$  functions which was implicitly assumed in  $(\mathcal{H}3)$  and  $(\mathcal{H}4)$ . The latter assumptions are expressed in a reduced form. One can check that the following conditions on the coefficients of the original SDE are sufficient for them to hold :

- $f$  and  $\sigma$  are bounded  $\mathcal{C}^4$  functions with bounded derivatives.
- $b$  is a bounded  $\mathcal{C}^3$  function with bounded derivatives.
- $\exists \sigma_0 > 0$  such that  $\forall y \in \mathbb{R}, \sigma(y) \geq \sigma_0$ .

*Proof of the theorem :* Throughout the proof, we denote by  $C$  a constant which can change from one line to another while always being independent of  $N$ . Thanks to Lemma 3, we just have to control the error on  $\tilde{X}$  :

$$\begin{aligned} \mathbb{E} \left[ \max_{0 \leq k \leq N} |\tilde{X}_{t_k} - \tilde{X}_{t_k}^N|^{2p} \right] &= \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \rho(F(Y_{t_k}) - F(\tilde{Y}_{t_k}^N)) + \sum_{j=0}^{k-1} \left( \int_{t_j}^{t_{j+1}} h(Y_s) ds - \delta_N h(\tilde{Y}_{t_j}^N) \right. \right. \right. \\ &\quad \left. \left. + \sqrt{\frac{1-\rho^2}{\delta_N}} \int_{t_j}^{t_{j+1}} \psi(Y_s) ds \Delta B_{j+1} \right. \right. \\ &\quad \left. \left. - \sqrt{1-\rho^2} \sqrt{\left( \psi(\tilde{Y}_{t_j}^N) + \frac{\sigma \psi'(\tilde{Y}_{t_j}^N)}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right) \vee \underline{\psi}} \Delta B_{j+1} \right|^{2p} \right] \\ &\leq 3^{2p-1} (\rho^{2p} I_0 + I_1 + (1-\rho^2)^p I_2) \end{aligned}$$

where

$$\begin{aligned} I_0 &= \mathbb{E} \left[ \max_{0 \leq k \leq N} |F(Y_{t_k}) - F(\tilde{Y}_{t_k}^N)|^{2p} \right] \\ I_1 &= \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \sum_{j=0}^{k-1} \left( \int_{t_j}^{t_{j+1}} h(Y_s) ds - \delta_N h(\tilde{Y}_{t_j}^N) \right) \right|^{2p} \right] \end{aligned}$$

and

$$\begin{aligned} I_2 &= \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \sum_{j=0}^{k-1} \left( \sqrt{\frac{1-\rho^2}{\delta_N}} \int_{t_j}^{t_{j+1}} \psi(Y_s) ds - \right. \right. \right. \\ &\quad \left. \left. \sqrt{\left( \psi(\tilde{Y}_{t_j}^N) + \frac{\sigma \psi'(\tilde{Y}_{t_j}^N)}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right) \vee \underline{\psi}} \Delta B_{j+1} \right|^{2p} \right]. \end{aligned}$$

$(\mathcal{H}1)$  yields that  $F$  is Lipschitz continuous so using Lemma 3 we show that  $I_0 \leq \frac{C}{N^{2p}}$ . Next, we have that

$$I_1 \leq C \left( \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \sum_{j=0}^{k-1} \int_{t_j}^{t_{j+1}} h(Y_s) ds - \delta_N h(Y_{t_j}) \right|^{2p} \right] + \delta_N^{2p} \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \sum_{j=0}^{k-1} h(Y_{t_j}) - h(\tilde{Y}_{t_j}^N) \right|^{2p} \right] \right).$$

On one hand, thanks to assumption  $(\mathcal{H}3)$  and Lemma 3,

$$\delta_N^{2p} \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \sum_{j=0}^{k-1} h(Y_{t_j}) - h(\tilde{Y}_{t_j}^N) \right|^{2p} \right] \leq C \delta_N \sum_{j=0}^{N-1} \mathbb{E} \left[ |h(Y_{t_j}) - h(\tilde{Y}_{t_j}^N)|^{2p} \right] \leq \frac{C}{N^{2p}}.$$

On the other hand, using an integration by parts formula,



$$\begin{aligned}
\bar{I}_1 &:= \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \sum_{j=0}^{k-1} \int_{t_j}^{t_{j+1}} h(Y_s) - h(Y_{t_j}) ds \right|^{2p} \right] \\
&= \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \sum_{j=0}^{k-1} \int_{t_j}^{t_{j+1}} (t_{j+1} - s) \left( (bh' + \frac{\sigma^2 h''}{2})(Y_s) ds + \sigma h'(Y_s) dW_s \right) \right|^{2p} \right] \\
&\leq 2^{2p-1} \left( \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \int_0^{t_k} (\tau_s - s) (bh' + \frac{\sigma^2 h''}{2})(Y_s) ds \right|^{2p} \right] + \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \int_0^{t_k} (\tau_s - s) \sigma h'(Y_s) dW_s \right|^{2p} \right] \right)
\end{aligned}$$

where we denoted by  $\tau_s$  the lowest discretization point greater than  $s$  :  $\tau_s = \lceil \frac{s}{\delta_N} \rceil \delta_N$ . Using Jensen's inequality for the first integral and the Burkholder-Davis-Gundy inequality for the second, we obtain

$$\begin{aligned}
\bar{I}_1 &\leq C \left( \mathbb{E} \left[ \max_{0 \leq k \leq N} t_k^{2p-1} \int_0^{t_k} (\tau_s - s)^{2p} \left| (bh' + \frac{\sigma^2 h''}{2})(Y_s) \right|^{2p} ds \right] \right. \\
&\quad \left. + \mathbb{E} \left[ \left( \int_0^T (\tau_s - s)^2 |\sigma h'(Y_s)|^2 ds \right)^p \right] \right) \\
&\leq \frac{C}{N^{2p}} \int_0^T \mathbb{E} \left[ \left| (bh' + \frac{\sigma^2 h''}{2})(Y_s) \right|^{2p} + |\sigma h'(Y_s)|^{2p} \right] ds.
\end{aligned}$$

Under the assumptions of Lemma 3,  $\sup_{0 \leq t \leq T} \mathbb{E}(|Y_s|^{2p}) < \infty$  (see Problem 3.15 p. 306 of Karatzas and Shreve [16] for example) so, with the help of assumption  $(\mathcal{H}3)$ , we conclude that  $\bar{I}_1 \leq \frac{C}{N^{2p}}$  and hence  $I_1 \leq \frac{C}{N^{2p}}$ . We now turn to the last term. Using the Burkholder-Davis-Gundy inequality, we get

$$\begin{aligned}
I_2 &\leq C \delta_N^p \mathbb{E} \left[ \left( \sum_{j=0}^{N-1} \left( \sqrt{\frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} \psi(Y_s) ds} - \sqrt{\left( \psi(\tilde{Y}_{t_j}^N) + \frac{\sigma \psi'(\tilde{Y}_{t_j}^N)}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right) \vee \underline{\psi}} \right)^2 \right)^p \right] \\
&\leq \delta_N \sum_{j=0}^{N-1} \mathbb{E} \left[ \left| \sqrt{\frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} \psi(Y_s) ds} - \sqrt{\left( \psi(\tilde{Y}_{t_j}^N) + \frac{\sigma \psi'(\tilde{Y}_{t_j}^N)}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right) \vee \underline{\psi}} \right|^{2p} \right].
\end{aligned} \tag{8}$$

Assumption  $(\mathcal{H}2)$  yields that the two terms appearing in the square root are bounded from below by  $\underline{\psi} > 0$  so we have that

$$\begin{aligned}
I_2 &\leq C \delta_N \sum_{j=0}^{N-1} \mathbb{E} \left[ \left| \frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} \psi(Y_s) ds - \left( \psi(\tilde{Y}_{t_j}^N) + \frac{\sigma \psi'(\tilde{Y}_{t_j}^N)}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right) \vee \underline{\psi} \right|^{2p} \right] \\
&\leq C N^{2p-1} \sum_{j=0}^{N-1} \mathbb{E} \left[ \left| \int_{t_j}^{t_{j+1}} \psi(Y_s) ds - \left( \psi(\tilde{Y}_{t_j}^N) \delta_N + \sigma \psi'(\tilde{Y}_{t_j}^N) \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right) \right|^{2p} \right] \\
&\leq C N^{2p-1} \sum_{j=0}^{N-1} (\bar{I}_2^j + \tilde{I}_2^j)
\end{aligned}$$

where

$$\bar{I}_2^j = \mathbb{E} \left[ \left| \int_{t_j}^{t_{j+1}} \psi(Y_s) ds - \left( \psi(Y_{t_j}) \delta_N + \sigma \psi'(Y_{t_j}) \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right) \right|^{2p} \right] \tag{9}$$

and

$$\tilde{I}_2^j = \mathbb{E} \left[ \left| \delta_N \left( \psi(Y_{t_j}) - \psi(\tilde{Y}_{t_j}^N) \right) + \left( \sigma\psi'(Y_{t_j}) - \sigma\psi'(\tilde{Y}_{t_j}^N) \right) \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right|^{2p} \right].$$

Again, integrating by parts yields that

$$\tilde{I}_2^j = \mathbb{E} \left[ \left| \int_{t_j}^{t_{j+1}} (t_{j+1} - s) \left( (\sigma\psi'(Y_s) - \sigma\psi'(Y_{t_j})) dW_s + (b\psi' + \frac{\sigma^2}{2}\psi'')(Y_s) ds \right) \right|^{2p} \right]$$

We control the stochastic integral term as follows

$$\begin{aligned} \mathbb{E} \left[ \left| \int_{t_j}^{t_{j+1}} (t_{j+1} - s) (\sigma\psi'(Y_s) - \sigma\psi'(Y_{t_j})) dW_s \right|^{2p} \right] &\leq C\delta_N^{p-1} \mathbb{E} \left[ \int_{t_j}^{t_{j+1}} (t_{j+1} - s)^{2p} |\sigma\psi'(Y_s) - \sigma\psi'(Y_{t_j})|^{2p} ds \right] \\ &\leq C\delta_N^{3p-1} \int_{t_j}^{t_{j+1}} \mathbb{E} \left[ |\sigma\psi'(Y_s) - \sigma\psi'(Y_{t_j})|^{2p} \right] ds \\ &\leq C\delta_N^{3p-1} \int_{t_j}^{t_{j+1}} \mathbb{E} \left[ |Y_s - Y_{t_j}|^{2p} \right] ds \\ &\leq C\delta_N^{3p-1} \int_{t_j}^{t_{j+1}} |s - t_j|^p ds \\ &\leq C\delta_N^{4p}. \end{aligned}$$

The third inequality is due to assumption (H4) and the fourth one is a standard result on the control of the moments of the increments of the solution of a SDE with Lipschitz continuous coefficients (see Problem 3.15 p. 306 of Karatzas and Shreve [16] for example).

We also control the other term thanks to assumption (H4) :

$$\begin{aligned} \mathbb{E} \left[ \left| \int_{t_j}^{t_{j+1}} (t_{j+1} - s) (b\psi' + \frac{\sigma^2}{2}\psi'')(Y_s) ds \right|^{2p} \right] &\leq \delta_N^{2p-1} \mathbb{E} \left[ \int_{t_j}^{t_{j+1}} (t_{j+1} - s)^{2p} |b\psi' + \frac{\sigma^2}{2}\psi''(Y_s)|^{2p} ds \right] \\ &\leq \delta_N^{4p-1} \int_{t_j}^{t_{j+1}} \mathbb{E} \left[ \left| b\psi' + \frac{\sigma^2}{2}\psi''(Y_s) \right|^{2p} \right] ds \\ &\leq C\delta_N^{4p}. \end{aligned}$$

Hence,  $\tilde{I}_2^j \leq \frac{C}{N^{4p}}$ . To conclude the proof of the theorem, it remains to show a similar result for  $\tilde{I}_2^j$  :

$$\begin{aligned} \tilde{I}_2^j &\leq 2^{2p-1} \mathbb{E} \left[ \left| \delta_N \left( \psi(Y_{t_j}) - \psi(\tilde{Y}_{t_j}^N) \right) \right|^{2p} + \left| \left( \sigma\psi'(Y_{t_j}) - \sigma\psi'(\tilde{Y}_{t_j}^N) \right) \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right|^{2p} \right] \\ &\leq C \left( \delta_N^{2p} \mathbb{E} \left[ |Y_{t_j} - \tilde{Y}_{t_j}^N|^{2p} \right] + \frac{\delta_N^{3p}}{3^p} \mathbb{E} \left[ |Y_{t_j} - \tilde{Y}_{t_j}^N|^{2p} \right] \right) \\ &\leq \frac{C}{N^{4p}}. \end{aligned}$$

The second inequality is due to the fact that  $\psi$  is Lipschitz continuous (thanks to assumption (H1)) for the first term and to the independence of  $(\sigma\psi'(Y_{t_j}) - \sigma\psi'(\tilde{Y}_{t_j}^N))$  and  $\int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds$  for the second term.  $\square$

**Remark 5** — Our scheme exhibits the same convergence properties as the Cruzeiro et al. [7] scheme. In addition to the fact that it involves fewer terms, it presents the advantage of improving the multilevel Monte Carlo convergence. This method, which is a generalization of the statistical Romberg extrapolation method of Kebaier [17], was introduced by Giles [9, 10].

Indeed, consider the discretization scheme with time step  $\delta_{2N} = \frac{T}{2N}$  :

$$\forall 0 \leq k \leq 2N-1, \tilde{X}_{\frac{(k+1)T}{2N}}^{2N} = \tilde{X}_{\frac{kT}{2N}}^{2N} + \rho \left( F(\tilde{Y}_{\frac{(k+1)T}{2N}}^{2N}) - F(\tilde{Y}_{\frac{kT}{2N}}^{2N}) \right) + \delta_{2N} h(\tilde{Y}_{\frac{kT}{2N}}^{2N}) + \sqrt{1-\rho^2} \\ \times \sqrt{\left( \psi(\tilde{Y}_{\frac{kT}{2N}}^{2N}) + \frac{\sigma \psi'(\tilde{Y}_{\frac{kT}{2N}}^{2N})}{\delta_{2N}} \int_{\frac{kT}{2N}}^{\frac{(k+1)T}{2N}} (W_s - W_{\frac{kT}{2N}}) ds \right) \vee \underline{\psi} \left( B_{\frac{(k+1)T}{2N}} - B_{\frac{kT}{2N}} \right)}.$$

Denote by  $v_k^{2N} = \sqrt{1-\rho^2} \sqrt{\left( \psi(\tilde{Y}_{\frac{kT}{2N}}^{2N}) + \frac{\sigma \psi'(\tilde{Y}_{\frac{kT}{2N}}^{2N})}{\delta_{2N}} \int_{\frac{kT}{2N}}^{\frac{(k+1)T}{2N}} (W_s - W_{\frac{kT}{2N}}) ds \right) \vee \underline{\psi}}$  the random variable which multiplies the increment of the Brownian motion  $(B_{\frac{(k+1)T}{2N}} - B_{\frac{kT}{2N}})$ . Because of the independence properties,  $(\tilde{X}_{t_k}^N)_{0 \leq k \leq N}$  has the same distribution law as the vector  $(\tilde{X}_{t_k}^N)_{0 \leq k \leq N}$  defined inductively by  $\tilde{X}_{t_0}^N = \log(s_0)$  and

$$\forall 0 \leq k \leq N-1, \tilde{X}_{t_{k+1}}^N = \tilde{X}_{t_k}^N + \rho \left( F(\tilde{Y}_{t_{k+1}}^N) - F(\tilde{Y}_{t_k}^N) \right) + \delta_N h(\tilde{Y}_{t_k}^N) \\ + \sqrt{1-\rho^2} \sqrt{\left( \psi(\tilde{Y}_{t_k}^N) + \frac{\sigma \psi'(\tilde{Y}_{t_k}^N)}{\delta_N} \int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds \right) \vee \underline{\psi}} \Delta \tilde{B}_{k+1}^N$$

where

$$\Delta \tilde{B}_{k+1}^N = \sqrt{2} \left[ \frac{v_{2k}^{2N} \left( B_{\frac{(2k+1)T}{2N}} - B_{\frac{2kT}{2N}} \right) + v_{2k+1}^{2N} \left( B_{\frac{(2k+2)T}{2N}} - B_{\frac{(2k+1)T}{2N}} \right)}{\sqrt{(v_{2k}^{2N})^2 + (v_{2k+1}^{2N})^2}} \right]. \quad (10)$$

Going over the proof of the theorem, one can show in the same way that

$$\mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \tilde{X}_{t_k}^N - \tilde{X}_{t_k}^{2N} \right|^2 \right] = \mathcal{O}(N^{-2}). \quad (11)$$

Hence, one can apply the multilevel Monte Carlo method to compute the expectation of a Lipschitz continuous functional of  $X$  and reduce the computational cost to achieve a desired root-mean-square error of  $\epsilon > 0$  to a  $\mathcal{O}(\epsilon^{-2})$ .

To summarize, the particular structure of our scheme enabled us to reconstruct the coupling which allows to efficiently control the error between the scheme with time step  $\frac{T}{N}$  and the one with time step  $\frac{T}{2N}$ . This does not seem possible with the Cruzeiro et al. [7] scheme.

From a practical point of view, it is more interesting to obtain a convergence result for the stock price. It is also more challenging because the exponential function is not globally Lipschitz continuous. We can nevertheless state the following corollary with some general assumptions and we will see in the next section that we can make them more precise in the case where  $(Y_t)_{t \in [0, T]}$  is an Ornstein-Uhlenbeck process.

**Corollary 6** — Let  $p \geq 1$ . Under the assumptions of Theorem 2 and if

$$(\mathcal{H}7) \quad \exists \epsilon > 0 \text{ such that } \mathbb{E} \left[ \max_{0 \leq k \leq N} S_{t_k}^{2p+\epsilon} \right] + \mathbb{E} \left[ \max_{0 \leq k \leq N} e^{(2p+\epsilon)\tilde{X}_{t_k}^N} \right] < \infty$$

then there exists a positive constant  $C$  independent of  $N$  such that

$$\mathbb{E} \left[ \max_{0 \leq k \leq N} \left| e^{\tilde{X}_{t_k}} - e^{\tilde{X}_{t_k}^N} \right|^{2p} \right] \leq \frac{C}{N^{2p}}.$$

*Proof :* Using Hölder inequality we have that

$$\begin{aligned} \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| e^{\tilde{X}_{t_k}} - e^{\tilde{X}_{t_k}^N} \right|^{2p} \right] &\leq \mathbb{E} \left[ \max_{0 \leq k \leq N} \left( e^{2p\tilde{X}_{t_k}} \vee e^{2p\tilde{X}_{t_k}^N} \right) \left| \tilde{X}_{t_k} - \tilde{X}_{t_k}^N \right|^{2p} \right] \\ &\leq \left( \mathbb{E} \left[ \max_{0 \leq k \leq N} S_{t_k}^{2p+\epsilon} \right] + \mathbb{E} \left[ \max_{0 \leq k \leq N} e^{(2p+\epsilon)\tilde{X}_{t_k}^N} \right] \right)^{\frac{2p}{2p+\epsilon}} \\ &\quad \times \left( \mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \tilde{X}_{t_k} - \tilde{X}_{t_k}^N \right|^{\frac{2p\epsilon+4p^2}{\epsilon}} \right] \right)^{\frac{\epsilon}{2p+\epsilon}}. \end{aligned}$$

We conclude by assumption  $(\mathcal{H}7)$  and Theorem 2.  $\square$

**Remark 7** — Had we introduced a new cut-off to our scheme as follows

$$\begin{aligned} \tilde{X}_{t_{k+1}}^N &= \tilde{X}_{t_k}^N + \rho \left( F(\tilde{Y}_{t_{k+1}}^N) - F(\tilde{Y}_{t_k}^N) \right) + \delta_N h(\tilde{Y}_{t_k}^N) \\ &\quad + \sqrt{1 - \rho^2} \sqrt{\left( \psi(\tilde{Y}_{t_k}^N) + \frac{\sigma \psi'(\tilde{Y}_{t_k}^N)}{\delta_N} \int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds \right) \wedge \bar{\psi} \vee \underline{\psi}} \Delta B_{k+1} \end{aligned}$$

assumption  $(\mathcal{H}7)$  would have been induced by assuming that the functions  $F, f$  and  $h$  are bounded.

## 1.2 Special case of an Ornstein-Uhlenbeck process driving the volatility

For many stochastic volatility models, the process  $(Y_t)_{t \in [0, T]}$  which drives the volatility is an Ornstein-Uhlenbeck process. For example, this is the case for all the models cited in the introduction but the Heston model. Therefore, it is useful to focus on this particular case. We will hereafter suppose that  $(Y_t)_{t \in [0, T]}$  is the solution of the following SDE

$$dY_t = \nu dW_t + \kappa(\theta - Y_t)dt, \quad Y_0 = y_0 \quad (12)$$

with  $\nu > 0$  and  $\kappa, \theta \in \mathbb{R}$ . Since exact simulation is possible, we can replace the Milstein discretization by the true solution in our previous scheme :

**WeakTraj\_1 scheme when  $Y$  is an O-U process**

$$\begin{aligned} \tilde{X}_{t_{k+1}}^N &= \tilde{X}_{t_k}^N + \rho \left( F(Y_{t_{k+1}}) - F(Y_{t_k}) \right) + \delta_N h(Y_{t_k}) \\ &\quad + \sqrt{1 - \rho^2} \sqrt{\left( \psi(Y_{t_k}) + \frac{\nu \psi'(Y_{t_k})}{\delta_N} \int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds \right) \vee \underline{\psi}} \Delta B_{k+1}. \end{aligned} \quad (13)$$

Note that we require the exact simulation of both  $(Y_{t_k}, Y_{t_{k+1}})$  and  $\int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds$ . The unique solution of (12) is  $Y_t = y_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t}) + \nu \int_0^t e^{-\kappa(t-s)} dW_s$  and one can easily deduce that,  $\forall k \in \{0, \dots, N-1\}$ ,

$$\begin{pmatrix} Y_{t_{k+1}} - e^{-\kappa \delta_N} Y_{t_k} \\ \int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds \end{pmatrix} \sim \mathcal{N}(M, \Gamma)$$

$$\text{where } M = \begin{pmatrix} \theta(1 - e^{-\kappa \delta_N}) \\ 0 \end{pmatrix} \text{ and } \Gamma = \begin{pmatrix} \frac{\nu^2}{2\kappa}(1 - e^{-2\kappa \delta_N}) & \frac{\nu}{\kappa^2}(1 - e^{-\kappa \delta_N}(1 + \kappa \delta_N)) \\ \frac{\nu}{\kappa^2}(1 - e^{-\kappa \delta_N}(1 + \kappa \delta_N)) & \frac{\delta_N^3}{3} \end{pmatrix}.$$

We first state the following technical lemma whose proof is postponed to the appendix :

**Lemma 8** —  $\forall c_1 > 0, c_2 \in [0, 1)$ ,

$$\mathbb{E} \left( e^{c_1 \sup_{0 \leq t \leq T} |Y_t|^{1+c_2}} \right) < \infty.$$

$$\text{Moreover, when } y_0 \neq 0, \forall \alpha > 0, \exists C < +\infty, \forall N \in \mathbb{N}^*, \sup_{t \in [0, T]} \mathbb{P} [|Y_t| \leq N^{-\alpha}] \leq C N^{-\alpha}. \quad (14)$$

As might be expected, it is possible to weaken the assumptions of Theorem 2. In particular, we relax the assumption on the lower bound of the volatility ( $\mathcal{H}2$ ) and replace it with a weaker one (see assumption ( $\mathcal{H}10$ ) below). The following theorem applies for Scott's model [27] (and therefore for the Hull and White [13] model) where we have  $h(y) = r - \frac{\sigma_0^2 e^{2y}}{2} - \rho \sigma_0 e^y (\frac{\kappa}{\nu}(\theta - y) + \frac{\nu}{2})$  and  $\psi(y) = \sigma_0^2 e^{2y}$ .

**Theorem 9** — Let  $p \geq 1$ . Suppose that  $Y$  is solution of (12) and that the scheme is defined by (13). Under assumption ( $\mathcal{H}2$ ) of Theorem 2 and if

( $\mathcal{H}8$ )  $f$  is a  $C^3$  function

( $\mathcal{H}9$ ) there exist three constants  $c_0 > 0, c_1 > 0$  and  $c_2 \in [0, 1)$  such that,  $\forall y \in \mathbb{R}$ ,

$$\begin{aligned} \left| \kappa(\theta - y)h'(y) + \frac{\nu^2}{2}h''(y) \right| &\leq c_0 e^{c_1 |y|^{1+c_2}} \\ |h'(y)| &\leq c_0 e^{c_1 |y|^{1+c_2}} \\ \left| \kappa(\theta - y)\psi'(y) + \frac{\nu^2}{2}\psi''(y) \right| &\leq c_0 e^{c_1 |y|^{1+c_2}} \\ |\psi''(y)| &\leq c_0 e^{c_1 |y|^{1+c_2}} \end{aligned}$$

then there exists a constant  $C$  independent of the number of time steps  $N$  such that

$$\mathbb{E} \left[ \max_{0 \leq k \leq N} |\tilde{X}_{t_k} - \tilde{X}_{t_k}^N|^{2p} \right] \leq \frac{C}{N^{2p}}$$

The same result holds true when we replace assumption ( $\mathcal{H}2$ ) by

( $\mathcal{H}10$ )

$$\forall y \in \mathbb{R}, \psi(y) > 0$$

$$\sup_{t \leq T} \mathbb{E} \left( \frac{1}{\psi^{p(1+\epsilon)}(Y_t)} \right) < \infty.$$

*Proof :* The proof of the first part of the theorem repeats the proof of Theorem 2 with fewer terms to control because of the exact simulation of  $(Y_t)_{t \in [0, T]}$ . At the places where we used assumptions  $(\mathcal{H}3)$  and  $(\mathcal{H}4)$ , we use assumption  $(\mathcal{H}9)$  together with Lemma 8.

We now focus on the second part of the theorem. According to equation (8), all we have to show is the existence of a positive constant  $C$  independent of  $N$  such that  $\forall j \in \{0, \dots, N-1\}$

$$\mathbb{E} \left[ \left| \sqrt{A_j} - \sqrt{D_j} \right|^{2p} \right] \leq \frac{C}{N^{2p}}$$

where

$$\begin{aligned} - A_j &= \frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} \psi(Y_s) ds, \\ - D_j &= \left( \psi(Y_{t_j}) + \frac{\nu \psi'(Y_{t_j})}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right) \vee \underline{\psi}. \end{aligned}$$

One has

$$|\sqrt{A_j} - \sqrt{D_j}|^{2p} = \frac{|A_j - D_j|^{2p}}{(\sqrt{A_j} + \sqrt{D_j})^{2p}} \leq \frac{|A_j - D_j|^{2p}}{A_j^p}.$$

Hence by Hölder's and Jensen's inequalities,

$$\begin{aligned} \mathbb{E} \left[ |\sqrt{A_j} - \sqrt{D_j}|^{2p} \right] &\leq \left( \mathbb{E} \left[ \frac{1}{A_j^{p(1+\epsilon)}} \right] \right)^{\frac{1}{1+\epsilon}} \left( \mathbb{E} \left[ |A_j - D_j|^{2p \frac{1+\epsilon}{\epsilon}} \right] \right)^{\frac{\epsilon}{1+\epsilon}} \\ &\leq \left( \frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} \mathbb{E} \left( \frac{1}{\psi^{p(1+\epsilon)}(Y_s)} \right) ds \right)^{\frac{1}{1+\epsilon}} \left( \mathbb{E} \left[ |A_j - D_j|^{2p \frac{1+\epsilon}{\epsilon}} \right] \right)^{\frac{\epsilon}{1+\epsilon}}. \end{aligned}$$

Thanks to assumption  $(\mathcal{H}10)$ , the first term in the right-hand-side is smaller than a finite constant not depending on  $N$  whereas the second term is smaller than  $C \delta_N^{2p}$  according to the estimation of  $\bar{I}_2^j$  in the proof of Theorem 2.  $\square$

The following proposition is dedicated to the Stein and Stein [28] and the quadratic Gaussian models which satisfy neither assumption  $(\mathcal{H}2)$  nor assumption  $(\mathcal{H}10)$  since  $\psi(y)$  vanishes at the origin.

**Proposition 10** — *Suppose that  $Y$  is solution of (12) starting from  $y_0$  and that the scheme is defined by (13). Then for  $p \geq 1$  and  $\varepsilon > 0$ , there is a constant  $C$  not depending on  $N$  such that*

$$\mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \tilde{X}_{t_k} - \tilde{X}_{t_k}^N \right|^{2p} \right] \leq \begin{cases} \frac{C}{N^{\frac{2p+1}{2}-\varepsilon}} & \text{in the Stein and Stein model when } y_0 \neq 0 \\ \frac{C}{N^{2p-\varepsilon}} & \text{in the quadratic Gaussian model} \end{cases}.$$

*In particular, the order of convergence in  $L^2$  is not smaller than  $3/4 - \varepsilon$  in the Stein and Stein model.*

*Proof :* In both the Stein and Stein and the quadratic Gaussian models the function  $h(y)$ , respectively equal to  $r - \frac{y^2}{2} - \rho \left( \frac{\kappa y}{\nu} (\theta - y) + \frac{\nu}{2} \right)$  and  $r - \frac{y^4}{2} - \rho \left( \frac{\kappa y^2}{\nu} (\theta - y) + \nu y \right)$  satisfies the first two inequalities of assumption  $(\mathcal{H}9)$ . So we only need to focus on  $\mathbb{E} \left[ \left| \sqrt{A_j} - \sqrt{D_j} \right|^{2p} \right]$  with

$$(A_j, D_j, \psi(y)) = \begin{cases} \left( \frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} Y_s^2 ds, \left( Y_{t_j}^2 + \frac{2\nu Y_{t_j}}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right)^+, y^2 \right) & \text{in the Stein and Stein model} \\ \left( \frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} Y_s^4 ds, \left( Y_{t_j}^4 + \frac{4\nu Y_{t_j}^3}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right)^+, y^4 \right) & \text{in the quadratic Gaussian model} \end{cases}.$$

So in both cases  $D_j \leq \frac{\psi(Y_{t_j})}{2} \Rightarrow \left| \frac{4\nu}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right| \geq \frac{|Y_{t_j}|}{2}$ . In what follows, we use repeatedly that the power function  $\psi$  is non-decreasing on the positive half line and commutes with other power functions. Let  $(\alpha_l)_{1 \leq l \leq L}$  be a decreasing sequence in  $(0, \frac{1}{2})$ . Using the convention  $N^{-\alpha_{L+1}} = +\infty$ , one has

$$\begin{aligned}
\mathbb{E} \left[ \left| \sqrt{A_j} - \sqrt{D_j} \right|^{2p} \right] &\leq \mathbb{E} \left[ \left| \sqrt{A_j} - \sqrt{D_j} \right|^{2p} \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_1}\}} \right] \\
&\quad + \sum_{l=1}^L \mathbb{E} \left[ \frac{|A_j - D_j|^{2p}}{(\sqrt{A_j} + \sqrt{D_j})^{2p}} \mathbb{1}_{\{N^{-\alpha_l} \leq |Y_{t_j}| < N^{-\alpha_{l+1}}, D_j \geq \frac{\psi(Y_{t_j})}{2}\}} \right] \\
&\quad + \sum_{l=1}^L \mathbb{E} \left[ \left| \sqrt{A_j} - \sqrt{D_j} \right|^{2p} \mathbb{1}_{\{N^{-\alpha_l} \leq |Y_{t_j}| < N^{-\alpha_{l+1}}, \left| \frac{4\nu}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right| \geq \frac{|Y_{t_j}|}{2}\}} \right] \\
&\leq \mathbb{E} \left[ (A_j^p + D_j^p) \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_1}\}} \right] + C \sum_{l=1}^L \psi(N^{p\alpha_l}) \mathbb{E} \left[ |A_j - D_j|^{2p} \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_{l+1}}\}} \right] \\
&\quad + \mathbb{E} \left[ (A_j^p + D_j^p) \mathbb{1}_{\{\left| \frac{4\nu}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right| > \frac{N^{-\alpha_1}}{2}\}} \right].
\end{aligned} \tag{15}$$

Since conditionally on  $Y_{t_j}$ , for  $s \in [t_j, t_{j+1}]$ ,  $Y_s \sim \mathcal{N}_1(Y_{t_j} e^{-\kappa(s-t_j)} + \theta(1 - e^{-\kappa(s-t_j)}), \frac{\nu^2}{2\kappa}(1 - e^{-2\kappa(s-t_j)}))$  (convention :  $\frac{\nu^2}{2\kappa}(1 - e^{-2\kappa(s-t_j)}) = \nu^2(s - t_j)$  when  $\kappa = 0$ ) and  $\frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \sim \mathcal{N}_1(0, \frac{\delta_N}{3})$  is independent from  $Y_{t_j}$  the first term in the right-hand-side of (15) is not greater than

$$\begin{aligned}
&\mathbb{E} \left[ \frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} \mathbb{E} [\psi(Y_s)^p | Y_{t_j}] ds \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_1}\}} \right] + 2^{p-1} \mathbb{E} \left[ \left( \psi(Y_{t_j})^p + \left| \frac{\nu\psi'(Y_{t_j})}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right|^p \right) \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_1}\}} \right] \\
&\leq C \mathbb{E} \left[ (\psi(Y_{t_j})^p + \psi(\delta_N^{\frac{p}{2}})) \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_1}\}} \right] + C(\psi(N^{-p\alpha_1}) + \psi'(N^{-p\alpha_1}) N^{-\frac{p}{2}}) \mathbb{P}(|Y_{t_j}| < N^{-\alpha_1}) \\
&\leq C\psi(N^{-p\alpha_1}) \mathbb{P}(|Y_{t_j}| < N^{-\alpha_1}),
\end{aligned}$$

where we used  $\alpha_1 < \frac{1}{2}$  for the last inequality. By Hölder's inequality and since  $\alpha_1 < \frac{1}{2}$  and  $G = \sqrt{\frac{3}{\delta_N^3}} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \sim \mathcal{N}_1(0, 1)$ , the third term in the right-hand-side of (15) is not greater than a constant multiplying

$$\mathbb{E}^{\frac{1}{1+\xi}} \left[ \frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} \psi(Y_s)^{p(1+\xi)} ds + \psi(Y_{t_j})^{p(1+\xi)} + \frac{|\psi'(Y_{t_j})|^{p(1+\xi)}}{\delta_N} \int_{t_j}^{t_{j+1}} \psi((s - t_j)^{\frac{p(1+\xi)}{4}}) ds \right] \mathbb{P}^{\frac{\xi}{1+\xi}} \left( |G| \geq CN^{\frac{1}{2}-\alpha_1} \right).$$

The first term of the product is bounded whereas, by the usual bound of the tail of the normal law  $\forall t > 0, \mathbb{P}(|G| \geq t) \leq \frac{2e^{-\frac{t^2}{2}}}{t\sqrt{2\pi}}$ , the product of the second term by any polynomial function of  $N$  is bounded.

Now for  $l \in \{1, \dots, L\}$ ,  $\mathbb{E} \left[ |A_j - D_j|^{2p} \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_{l+1}}\}} \right]$  is not greater than

$$\frac{1}{\delta_N^{2p}} \mathbb{E} \left[ \left| \int_{t_j}^{t_{j+1}} (t_{j+1} - s) \left( (\nu\psi'(Y_s) - \nu\psi'(Y_{t_j})) dW_s + ((b\psi' + \frac{\nu^2}{2}\psi'')(Y_s)) ds \right) \right|^{2p} \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_{l+1}}\}} \right],$$

with  $b(y) = \kappa(\theta - y)$ . As, by convention,  $N^{-\alpha_{L+1}} = +\infty$ , reasoning like in the estimation of  $\bar{I}_2^j$  in the proof of Theorem 2, one checks that the term with index  $l = L$  in the sum in the right-hand-side of (15) is smaller than  $C\psi(N^{p\alpha_L})N^{-2p}$ . In the quadratic Gaussian model, since  $\psi'(y) = 3y^3$ , using Burkholder-Davis-Gundy

inequality then Hölder's inequality, one obtains that for  $\xi > 0$ ,

$$\begin{aligned}
& \mathbb{E} \left[ \left| \int_{t_j}^{t_{j+1}} (t_{j+1} - s) (\psi'(Y_s) - \psi'(Y_{t_j})) dW_s \right|^{2p} \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_{l+1}}\}} \right] \\
& \leq \frac{C}{N^{p-1}} \int_{t_j}^{t_{j+1}} (t_{j+1} - s)^{2p} \mathbb{E} \left[ |Y_s - Y_{t_j}|^{2p} (|Y_s|^{4p} + |Y_{t_j}|^{4p}) \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_{l+1}}\}} \right] ds \\
& \leq \frac{C}{N^{3p-1}} \int_{t_j}^{t_{j+1}} \mathbb{E}^{\frac{\xi}{1+\xi}} \left[ |Y_s - Y_{t_j}|^{\frac{2p(1+\xi)}{\xi}} \right] \mathbb{E}^{\frac{1}{1+\xi}} \left[ (|Y_s|^{4p(1+\xi)} + |Y_{t_j}|^{4p(1+\xi)}) \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_{l+1}}\}} \right] ds \\
& \leq \frac{C}{N^{3p-1}} \int_{t_j}^{t_{j+1}} \frac{C}{N^p} \times \frac{C \mathbb{P}^{\frac{1}{1+\xi}}(|Y_{t_j}| < N^{-\alpha_{l+1}})}{N^{4p\alpha_{l+1}}} ds \leq C \frac{\mathbb{P}^{\frac{1}{1+\xi}}(|Y_{t_j}| < N^{-\alpha_{l+1}})}{N^{4p(1+\alpha_{l+1})}}.
\end{aligned}$$

In the Stein and Stein model, since  $\psi'(y) = 2y$ , one can only take advantage of the indicator function in the probability in the numerator and the power of  $N$  in the denominator is reduced to  $4p$ . In both models, the same bound with  $\xi = 0$  can be derived for  $\mathbb{E} \left[ \left| \int_{t_j}^{t_{j+1}} (t_{j+1} - s) ((b\psi' + \frac{\nu^2}{2}\psi'')(Y_s)) ds \right|^{2p} \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_{l+1}}\}} \right]$  and one concludes that

$$\mathbb{E} \left[ |A_j - D_j|^{2p} \mathbb{1}_{\{|Y_{t_j}| < N^{-\alpha_{l+1}}\}} \right] \leq \begin{cases} C \frac{\mathbb{P}^{\frac{1}{1+\xi}}(|Y_{t_j}| < N^{-\alpha_{l+1}})}{N^{2p}} & \text{in the Stein and Stein model} \\ C \frac{\mathbb{P}^{\frac{1}{1+\xi}}(|Y_{t_j}| < N^{-\alpha_{l+1}})}{N^{2p(1+2\alpha_{l+1})}} & \text{in the quadratic Gaussian model} \end{cases}.$$

Plugging the three estimations together with (14) in (15), one deduces that in the Stein and Stein model, when  $y_0 \neq 0$ ,

$$\mathbb{E} \left[ \left| \sqrt{A_j} - \sqrt{D_j} \right|^{2p} \right] \leq C \left( N^{-(2p+1)\alpha_1} + \sum_{l=1}^{L-1} N^{-2p(1-\alpha_l) - \frac{\alpha_{l+1}}{1+\xi}} + N^{-2p(1-\alpha_L)} \right).$$

Now we may suppose that  $\varepsilon < \frac{1}{2}$  since the smaller  $\varepsilon$  is the stronger the statement of the proposition is. We choose  $\alpha_l = \frac{1}{2} - \frac{l\varepsilon}{2p+1}$  for  $l \in \{1, \dots, L-1\}$  with  $L = \lceil \frac{(2p+1)(1-2\varepsilon)}{4p\varepsilon} \rceil$  and  $\alpha_L = \frac{2p-1+2\varepsilon}{4p}$  and  $\xi = \frac{2\varepsilon}{1-2\varepsilon}$ . Then  $(2p+1)\alpha_1 = 2p(1-\alpha_L) = \frac{2p+1}{2} - \varepsilon$  and for  $l \in \{1, \dots, L-1\}$ ,  $2p(1-\alpha_l) + \frac{\alpha_{l+1}}{1+\xi} \geq \frac{2p+1}{2} + (2pl - \frac{l+1}{1+\xi})\frac{\varepsilon}{2p+1} - \frac{\xi}{2(1+\xi)} \geq \frac{2p+1}{2} + 0 - \varepsilon$ .

In the quadratic Gaussian model, plugging the three estimations in (15), one obtains that

$$\mathbb{E} \left[ \left| \sqrt{A_j} - \sqrt{D_j} \right|^{2p} \right] \leq C \left( N^{-(4p+1\mathbb{1}_{\{y_0 \neq 0\}})\alpha_1} + \sum_{l=1}^{L-1} N^{-2p(1+2(\alpha_{l+1}-\alpha_l)) - \mathbb{1}_{\{y_0 \neq 0\}} \frac{\alpha_{l+1}}{1+\xi}} + N^{-2p(1-2\alpha_L)} \right). \quad (16)$$

We choose  $\alpha_l = \frac{2p-l\varepsilon}{4p}$  for  $l \in \{1, \dots, L-1\}$  with  $L = \lceil \frac{2p}{\varepsilon} \rceil - 1$  and  $\alpha_L = \frac{\varepsilon}{4p}$ . Then  $4p\alpha_1 = 2p(1-2\alpha_L) = 2p - \varepsilon = 2p(1+2(\alpha_{l+1}-\alpha_l))$  for  $l \in \{1, \dots, L-2\}$  and  $2p(1+2(\alpha_L - \alpha_{L-1})) = (\lceil \frac{2p}{\varepsilon} \rceil - 1)\varepsilon \geq 2p - \varepsilon$ .  $\square$

**Remark 11** —

- In the quadratic Gaussian model, when  $y_0 \neq 0$ , choosing  $\xi = 1$ ,  $\alpha_l = \frac{2p}{4p+1} \left( \frac{8p}{8p+1} \right)^{l-1}$  for  $l \in \{1, \dots, L\}$  with  $L = \lceil \frac{2 \log \log N}{\log \frac{8p+1}{8p}} \rceil$  in (16), one obtains that  $\mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \tilde{X}_{t_k} - \tilde{X}_{t_k}^N \right|^{2p} \right] \leq \frac{C \log \log N}{N^{2p}}$ .



- The fact that we can simulate exactly the volatility process without affecting the order of convergence of the scheme is yet another advantage of our approach over the Cruzeiro et al. [7] scheme. On the other hand, the Kahl and Jäckel [14] scheme allows the exact simulation of  $(Y_t)_{t \in [0, T]}$ . Applied to the SDE (3), it writes as

$$X_{t_{k+1}}^{IJK} = X_{t_k}^{IJK} + \left( r - \frac{f^2(Y_{t_{k+1}}) + f^2(Y_{t_k})}{4} \right) \delta_N + \rho f(Y_{t_k}) \Delta W_{k+1} + \sqrt{1 - \rho^2} \frac{f(Y_{t_{k+1}}) + f(Y_{t_k})}{2} \Delta B_{k+1} + \frac{\rho \nu}{2} f'(Y_{t_k}) \left( (\Delta W_{k+1})^2 - \delta_N \right). \quad (17)$$

Note that it is close to our scheme insofar as it takes advantage of the structure of the SDE (for example, unlike the Cruzeiro et al. [7] scheme, it allows the use of the coupling introduced in Remark 5). The main difference, which explains why our scheme has better weak trajectorial convergence order, is that we discretize more accurately the integral of  $f(Y_t)$  with respect to the Brownian motion  $(B_t)_{t \in [0, T]}$ . If, instead of a trapezoidal method, one uses the same discretization as for the WeakTraj\_1 scheme, then it can be shown that this modified IJK scheme will exhibit a first order weak trajectorial convergence.

- It is possible to improve the convergence at fixed times up to the order  $\frac{3}{2}$ . Following Lapeyre and Temam [20] who approximate an integral of the form  $\int_{t_k}^{t_{k+1}} g(Y_s) ds$  for a twice differentiable function  $g$  by  $\delta_N g(Y_{t_k}) + \nu g'(Y_{t_k}) \int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds + (\kappa(\theta - Y_{t_k}) g'(Y_{t_k}) + \frac{\nu^2}{2} g''(Y_{t_k})) \frac{\delta_N^2}{2}$ , we obtain the following scheme

**OU-Improved scheme**

$$\tilde{X}_{t_{k+1}}^N = \tilde{X}_{t_k}^N + \rho (F(Y_{t_{k+1}}) - F(Y_{t_k})) + \tilde{h}_k + \sqrt{1 - \rho^2} \sqrt{\tilde{\psi}_k} \Delta B_{k+1}, \quad (18)$$

where  $\tilde{h}_k = \delta_N h(Y_{t_k}) + \nu h'(Y_{t_k}) \int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds + (\kappa(\theta - Y_{t_k}) h'(Y_{t_k}) + \frac{\nu^2}{2} h''(Y_{t_k})) \frac{\delta_N^2}{2}$  and  $\tilde{\psi}_k = \left( \psi(Y_{t_k}) + \frac{\nu \psi'(Y_{t_k})}{\delta_N} \int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds + (\kappa(\theta - Y_{t_k}) \psi'(Y_{t_k}) + \frac{\nu^2}{2} \psi''(Y_{t_k})) \frac{\delta_N}{2} \right) \vee \underline{\psi}$ .

Mimicking the proof of Theorem 2, one can show that

$$\max_{0 \leq k \leq N} \mathbb{E} \left[ \left| \hat{X}_{t_k} - \hat{X}_{t_{k+1}}^N \right|^2 \right] = \mathcal{O}(N^{-3}),$$

where  $\hat{X}_{t_k}$  and  $\hat{X}_{t_{k+1}}^N$  have respectively the same distribution as  $X_{t_k}$  and  $\tilde{X}_{t_k}^N$ :

$$\hat{X}_{t_k} = X_0 + \rho (F(Y_{t_k}) - F(y_0)) + \int_0^{t_k} h(Y_s) ds + \sqrt{1 - \rho^2} \sqrt{\frac{1}{t_k} \int_0^{t_k} \psi(Y_s) ds} B_{t_k}$$

and

$$\hat{X}_{t_k}^N = X_0 + \rho (F(Y_{t_k}) - F(y_0)) + \sum_{j=0}^{k-1} \tilde{h}_j + \sqrt{1 - \rho^2} \sqrt{\frac{\delta_N}{t_k} \sum_{j=0}^{k-1} \tilde{\psi}_j} B_{t_k}.$$

As for the stock, we can prove the same convergence result under some additional assumptions which are more explicit than assumption (H7) of Corollary 6. To do so, let us make the following changes in our scheme so that we can control its exponential moments:

$$\begin{aligned} \tilde{X}_{t_{k+1}}^N &= \tilde{X}_{t_k}^N + \rho (F(Y_{t_{k+1}}) - F(Y_{t_k})) + \delta_N h(Y_{t_k}) \\ &+ \sqrt{1 - \rho^2} \sqrt{\left( \psi(Y_{t_k}) + \frac{\nu \psi'(Y_{t_k})}{\delta_N} \int_{t_k}^{t_{k+1}} (W_s - W_{t_k}) ds \right) \wedge \hat{\psi}(Y_{t_k}) \vee \underline{\psi}} \Delta B_{k+1}. \end{aligned} \quad (19)$$

**Proposition 12** — Suppose that  $Y$  is solution of (12) and that the scheme is defined by (19). Under the assumptions (H8), (H9) and (H10) of Theorem 9 and if

(H11) there exists  $\beta \in (0, 1)$  and  $K > 0$  such that  $\forall y \in \mathbb{R}$

$$|h(y)| + |F(y)| \leq K(1 + |y|^{1+\beta})$$

$$|f(y)| \leq K(1 + |y|^\beta)$$

$$|f'(y)| \leq K|f(y)|$$

then,  $\forall p \geq 1$ , there exists a positive constant  $C$  independent of  $N$  such that

$$\mathbb{E} \left[ \max_{0 \leq k \leq N} |e^{\tilde{X}_{t_k}} - e^{\tilde{X}_{t_k}^N}|^{2p} \right] \leq \frac{C}{N^{2p}}.$$

The same result holds true if one replaces assumption (H10) by assumption (H2).

*Proof :* We go over the proof of Corollary 6. The fact that  $\mathbb{E} \left[ \max_{0 \leq k \leq N} |\tilde{X}_{t_k} - \tilde{X}_{t_k}^N|^{4p} \right] = \mathcal{O}(\frac{1}{N^{4p}})$  is not a straightforward consequence of Theorem 9 anymore because we have introduced some changes in our scheme. However, looking through the proof of the theorem, one can see that it is enough to prove the following inequality :  $\forall j \in \{0, \dots, N-1\}$

$$\mathbb{E} \left[ \left| \sqrt{\frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} \psi(Y_s) ds} - \sqrt{\left( \psi(Y_{t_j}) + \frac{\nu \psi'(Y_{t_j})}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right) \wedge \hat{\psi}(Y_{t_j}) \vee \underline{\psi}} \right|^{2p} \right] \leq \frac{C}{N^{2p}}. \quad (20)$$

When  $\bar{\psi}$  is finite, since  $\frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} \psi(Y_s) ds$  is smaller than  $\hat{\psi}(Y_{t_j}) = \bar{\psi}$ , the expectation is not greater than the similar one without the new cut-off and (20) holds by the proof of Theorem 9. When  $\bar{\psi} = +\infty$ , for  $\epsilon > 0$ , the expectation of interest is smaller than

$$\begin{aligned} & \mathbb{E} \left[ \left| \sqrt{\frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} \psi(Y_s) ds} - \sqrt{\left( \psi(Y_{t_j}) + \frac{\nu \psi'(Y_{t_j})}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \right) \vee \underline{\psi}} \right|^{2p} \right] \\ & + \mathbb{E}^{\frac{1}{1+\epsilon}} \left[ \left( \frac{1}{\delta_N} \int_{t_j}^{t_{j+1}} \psi(Y_s)^p ds + \left( \frac{3}{2} \psi(Y_{t_j}) \right)^p \right)^{1+\epsilon} \right] \mathbb{P}^{\frac{\epsilon}{1+\epsilon}} \left[ \psi(Y_{t_j}) + \frac{\nu \psi'(Y_{t_j})}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds \geq \frac{3}{2} \psi(Y_{t_j}) \right]. \end{aligned}$$

We estimate the first term like in the proof of Theorem 9. The expectation in the second term is bounded uniformly in  $N$ . By hypothesis (H11),  $\exists K < +\infty$ ,  $\forall y \in \mathbb{R}$ ,  $|\psi'(y)| \leq K\psi(y)$  and, for  $G$  normally distributed, the probability in this second term is smaller than  $\mathbb{P} \left( |G| \geq \frac{C}{\sqrt{\delta_N}} \right)$  which decreases quicker than polynomially to 0 as  $N \rightarrow \infty$ . Therefore (20) holds.

Now, what is left to prove is that assumption (H7) is satisfied. On the one hand, we have that

$$\begin{aligned} \mathbb{E} \left[ \max_{0 \leq k \leq N} S_{t_k}^{4p} \right] &= \mathbb{E} \left[ \max_{0 \leq k \leq N} \left( S_0 + \int_0^{t_k} r S_s ds + \int_0^{t_k} f(Y_s) S_s \left( \rho dW_s + \sqrt{1 - \rho^2} dB_s \right) \right)^{4p} \right] \\ &\leq C \left( 1 + \int_0^T \mathbb{E} \left( S_t^{4p} (1 + f^{4p}(Y_t)) \right) dt \right) \\ &\leq C \left( 1 + \int_0^T \sqrt{\mathbb{E}(S_t^{8p})} \sqrt{\mathbb{E}((1 + f^{4p}(Y_t))^2)} dt \right). \end{aligned}$$

Thanks to assumption (H11) and Lemma 8, there exists  $C > 0$  such that  $\sqrt{\mathbb{E}((1 + f^{4p}(Y_t))^2)} \leq C$ . Observe that conditionally on  $(Y_t)_{t \in [0, T]}$ ,

$$X_t \sim \mathcal{N}\left(\log(s_0) + \rho(F(Y_t) - F(y_0)) + \int_0^t h(Y_s) ds, (1 - \rho^2) \int_0^t f^2(Y_s) ds\right), \quad (21)$$

so, by Jensen's inequality and assumption (H11)

$$\begin{aligned} \mathbb{E}(S_t^{8p}) &= \mathbb{E}\left(e^{8p(\log(s_0) + \rho(F(Y_t) - F(y_0)) + \int_0^t h(Y_s) ds)} e^{32p^2(1 - \rho^2) \int_0^t f^2(Y_s) ds}\right) \\ &\leq \mathbb{E}\left(e^{8p(\log(s_0) + \rho(F(Y_t) - F(y_0)))} \frac{1}{t} \int_0^t e^{t(8ph(Y_s) + 32p^2(1 - \rho^2)f^2(Y_s))} ds\right) \\ &\leq C \mathbb{E}\left(e^{C \sup_{0 \leq t \leq T} |Y_t|^{1+\beta}}\right). \end{aligned}$$

Using Lemma 8, we deduce that  $\mathbb{E}\left[\max_{0 \leq k \leq N} S_{t_k}^{4p}\right] < \infty$ .

On the other hand, using Cauchy-Schwarz inequality, we have that

$$\begin{aligned} \mathbb{E}\left[\max_{0 \leq k \leq N} e^{4p\tilde{X}_{t_k}^N}\right] &= \mathbb{E}\left[\max_{0 \leq k \leq N} \exp\left(4p\left(X_0 + \rho(F(Y_{t_k}) - F(y_0)) + \sum_{j=0}^{k-1} \delta_N h(Y_{t_j}) + \sum_{j=0}^{k-1} \sqrt{1 - \rho^2}\right.\right.\right. \\ &\quad \left.\left.\left.\times \sqrt{\left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds\right) \wedge \widehat{\psi}(Y_{t_j}) \vee \underline{\psi} \Delta B_{j+1}}\right)\right)\right] \\ &\leq \sqrt{\tilde{E}_1^N} \sqrt{\tilde{E}_2^N}, \end{aligned}$$

where

$$\tilde{E}_1^N = \mathbb{E}\left[\max_{0 \leq k \leq N} e^{8p(X_0 + \rho(F(Y_{t_k}) - F(y_0)) + \sum_{j=0}^{k-1} \delta_N h(Y_{t_j}))}\right]$$

and

$$\tilde{E}_2^N = \mathbb{E}\left[\max_{0 \leq k \leq N} e^{8p\sqrt{1 - \rho^2} \sum_{j=0}^{k-1} \sqrt{\left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds\right) \wedge \widehat{\psi}(Y_{t_j}) \vee \underline{\psi} \Delta B_{j+1}}}\right].$$

Using the same argument as before, we show that  $\tilde{E}_1^N \leq C \mathbb{E}\left(e^{C \sup_{0 \leq t \leq T} |Y_t|^{1+\beta}}\right) < \infty$ .

Denote by  $D_j = \left(\psi(Y_{t_j}) + \frac{\sigma\psi'(Y_{t_j})}{\delta_N} \int_{t_j}^{t_{j+1}} (W_s - W_{t_j}) ds\right) \wedge \widehat{\psi}(Y_{t_j}) \vee \underline{\psi}$ . Using Doob's maximal inequality for the positive submartingale  $\left(e^{4p\sqrt{1 - \rho^2} \sum_{j=0}^{k-1} \sqrt{D_j} \Delta B_{j+1}}\right)_{0 \leq k \leq N}$  (see Theorem 3.8 p. 13 of Karatzas and Shreve [16] for example), we also have that

$$\begin{aligned} \tilde{E}_2^N &\leq 4 \mathbb{E}\left(e^{8p\sqrt{1 - \rho^2} \sum_{j=0}^{N-1} \sqrt{D_j} \Delta B_{j+1}}\right) \\ &= 4 \mathbb{E}\left(\prod_{j=0}^{N-1} e^{32p^2 \delta_N (1 - \rho^2) D_j}\right) \\ &\leq 4 \mathbb{E}\left(\max_{0 \leq k \leq N-1} e^{32p^2 (1 - \rho^2) \widehat{\psi}(Y_{t_k})}\right). \end{aligned}$$

By virtue of assumption (H11),  $\tilde{E}_2^N < \infty$  which concludes the proof.  $\square$

## 2 A second order weak scheme

Integrating the first stochastic differential equation in (4) gives

$$X_t = \log(s_0) + \rho(F(Y_t) - F(y_0)) + \int_0^t h(Y_s)ds + \sqrt{1 - \rho^2} \int_0^t f(Y_s)dB_s. \quad (22)$$

We are only left with an integral with respect to time which can be handled by the use of a trapezoidal scheme and a stochastic integral where the integrand is independent of the Brownian motion. Hence, conditionally on  $(Y_t)_{t \in [0, T]}$ ,

$$X_T \sim \mathcal{N}(\log(s_0) + \rho(F(Y_T) - F(y_0)) + m_T, (1 - \rho^2)v_T), \quad (23)$$

where  $m_T = \int_0^T h(Y_s)ds$  and  $v_T = \int_0^T f^2(Y_s)ds$ . This suggests that, in order to properly approximate the law of  $X_T$ , one should accurately approximate the law of  $Y_T$  and carefully handle integrals with respect to time of functions of the process  $(Y_t)_{t \in [0, T]}$ . We thus define our weak scheme as follows

**Weak\_2 scheme**

$$\bar{X}_T^N = \log(s_0) + \rho(F(\bar{Y}_T^N) - F(y_0)) + \bar{m}_T^N + \sqrt{(1 - \rho^2)\bar{v}_T^N}G \quad (24)$$

where  $\bar{m}_T^N = \delta_N \sum_{k=0}^{N-1} \frac{h(\bar{Y}_{t_k}^N) + h(\bar{Y}_{t_{k+1}}^N)}{2}$ ,  $\bar{v}_T^N = \delta_N \sum_{k=0}^{N-1} \frac{f^2(\bar{Y}_{t_k}^N) + f^2(\bar{Y}_{t_{k+1}}^N)}{2}$ ,  $(\bar{Y}_{t_k}^N)_{0 \leq k \leq N}$  is the Ninomiya-Victoir scheme of  $(Y_t)_{t \in [0, T]}$  and  $G$  is an independent centered reduced Gaussian random variable. Note that, conditionally on  $(\bar{Y}_{t_k}^N)_{0 \leq k \leq N}$ ,  $\bar{X}_T^N$  is also a Gaussian random variable with mean  $\log(s_0) + \rho(F(\bar{Y}_T^N) - F(y_0)) + \bar{m}_T^N$  and variance  $(1 - \rho^2)\bar{v}_T^N$ .

It is well known that the Ninomiya and Victoir [25] scheme is of weak order two. For the sake of completeness, we give its definition in our setting :

$$\begin{cases} \bar{Y}_0^N = y_0 \\ \forall 0 \leq k \leq N-1, \bar{Y}_{t_{k+1}}^N = \exp\left(\frac{T}{2N}V_0\right) \exp\left((W_{t_{k+1}} - W_{t_k})V\right) \exp\left(\frac{T}{2N}V_0\right) (\bar{Y}_{t_k}^N), \end{cases}$$

where  $V_0 : x \mapsto b(x) - \frac{1}{2}\sigma\sigma'(x)$  and  $V : x \mapsto \sigma(x)$ . The notation  $\exp(tV)(x)$  stands for the solution, at time  $t$  and starting from  $x$ , of the ODE  $\eta'(t) = V(\eta(t))$ . What is nice with our setting is that we are in dimension one and thus such ODEs can be solved explicitly. Indeed, if  $\zeta$  is a primitive of  $\frac{1}{V} : \zeta(t) = \int_0^t \frac{1}{V(s)}ds$ , then the solution writes as  $\eta(t) = \zeta^{-1}(t + \zeta(x))$ .

Note that our scheme can be seen as a splitting scheme for the SDE satisfied by  $(Z_t = X_t - \rho F(Y_t), Y_t)$  :

$$\begin{cases} dZ_t &= h(Y_t)dt + \sqrt{1 - \rho^2}f(Y_t)dB_t \\ dY_t &= b(Y_t)dt + \sigma(Y_t)dW_t. \end{cases} \quad (25)$$

The differential operator associated to (25) writes as

$$\mathcal{L}v(z, y) = h(y)\frac{\partial v}{\partial z} + b(y)\frac{\partial v}{\partial y} + \frac{\sigma^2(y)}{2}\frac{\partial^2 v}{\partial y^2} + \frac{(1 - \rho^2)}{2}f^2(y)\frac{\partial^2 v}{\partial z^2} = \mathcal{L}_Y v(z, y) + \mathcal{L}_Z v(z, y),$$

where  $\mathcal{L}_Y v(z, y) = b(y)\frac{\partial v}{\partial y} + \frac{\sigma^2(y)}{2}\frac{\partial^2 v}{\partial y^2}$  and  $\mathcal{L}_Z v(z, y) = h(y)\frac{\partial v}{\partial z} + \frac{(1 - \rho^2)}{2}f^2(y)\frac{\partial^2 v}{\partial z^2}$ . One can check that our scheme amounts to first integrate exactly  $\mathcal{L}_Z$  over a half time step then apply the Ninomiya-Victoir scheme to  $\mathcal{L}_Y$  over a time step and finally integrate exactly  $\mathcal{L}_Z$  over a half time step. According to results on splitting (see Alfonsi [2] or Tanaka and Kohatsu-Higa [30] for example) one expects this scheme to exhibit second order weak convergence. Actually, according to Theorem 1.17 in Alfonsi [2], our scheme has potential second order of weak convergence. To deduce formally the order two of weak convergence, one only needs to check regularity of the solution of the backward Kolmogorov equation associated with the model.

**Remark 13** —

- As for plain vanilla options pricing, observe that, by the Romano and Touzi [26] formula,

$$\mathbb{E} \left( e^{-rT} \alpha(S_T) | (Y_t)_{t \in [0, T]} \right) = BS_{\alpha, T} \left( s_0 e^{\rho(F(Y_T) - F(y_0)) + m_T + \left( \frac{(1-\rho^2)v_T}{2T} - r \right) T}, \frac{(1-\rho^2)v_T}{T} \right),$$

where  $BS_{\alpha, T}(s, v)$  stands for the price of a European option with pay-off  $\alpha$  and maturity  $T$  in the Black & Scholes model with initial stock price  $s$ , volatility  $\sqrt{v}$  and constant interest rate  $r$ . When, like for a call or a put option,  $BS_{\alpha, T}$  is available in a closed form, one should approximate  $\mathbb{E} \left( e^{-rT} \alpha(S_T) \right)$  by

$$\frac{1}{M} \sum_{i=1}^M BS_{\alpha, T} \left( s_0 e^{\rho(F(\bar{Y}_T^{N, i}) - F(y_0)) + \bar{m}_T^{N, i} + \left( \frac{(1-\rho^2)\bar{v}_T^{N, i}}{2T} - r \right) T}, \frac{(1-\rho^2)\bar{v}_T^{N, i}}{T} \right),$$

where  $M$  is the total number of Monte Carlo samples and the index  $i$  refers to independent draws.

Indeed, the conditioning provides a variance reduction.

- In the special case of an Ornstein-Uhlenbeck process driving the volatility (i.e.  $(Y_t)_{t \in [0, T]}$  is solution of the SDE (12)), one should replace the Ninomiya-Victoir scheme by the true solution. The order two of weak convergence should then be preserved. Moreover, one can check that the OU-Improved scheme (18) has also potential second order of weak convergence. Better still, it achieves a weak trajectorial convergence of order  $\frac{3}{2}$  on the triplet  $(Y_t, m_t, v_t)_{t \in [0, T]}$  which allows for a significant improvement of the multilevel Monte Carlo method, as we shall check numerically.

### 3 Numerical comparative analysis of the proposed schemes with standard discretization methods

We focus on the case where  $(Y_t)_{t \in [0, T]}$  is an Ornstein-Uhlenbeck process since, as mentioned in the introduction, it encompasses several standard stochastic volatility models.

We are going to compare our schemes (WeakTraj.1, Weak.2 and OU-Improved) to the Euler scheme with exact simulation of the process  $(Y_t)_{t \in [0, T]}$  driving the volatility (hereafter denoted Euler), the Kahl and Jäckel [14] scheme (IJK) and the Cruzeiro *et al.* [7] scheme (CMT).

For the following numerical computations, unless otherwise stated, we are going to consider Scott's model (2). We use the same set of parameters as in Kahl and Jäckel [14] :  $S_0 = 100, r = 0.05, T = 1, \sigma_0 = 0.25, y_0 = 0, \kappa = 1, \theta = 0, \nu = \frac{7\sqrt{2}}{20}, \rho = -0.2$  and  $f : y \mapsto \sigma_0 e^y$ .

#### 3.1 Theoretical computational cost per timestep

The following table gives the computational cost per timestep for each scheme in terms of function evaluations and random samples needed. Except for the CMT scheme, we consider exact simulation of  $(Y_t)_{t \in [0, T]}$  which requires the simulation of one Gaussian variable at each timestep. For the Weak2 scheme, no other simulation per timestep is required since we only need to simulate one Gaussian variable at the terminal time (see equation (24)). The IJK scheme, as we can see according to (17), requires the simulation of the two Brownian increments.

Certainly, our schemes require more computational effort per timestep but we will see hereafter that their higher order of convergence suffices to have better efficiency.

Schemes	Function evaluations	Random number samples
WeakTraj_1	3 ( $f, f'$ and $F$ )	3 Gaussian simulations
Weak_2	2 ( $f$ and $f'$ )	1 Gaussian simulation
OU_Improved	5 ( $f, f', f'', f'''$ and $F$ )	3 Gaussian simulations
IJK	2 ( $f$ and $f'$ )	3 Gaussian simulations
CMT	2 ( $f$ and $f'$ )	2 Gaussian simulations
Euler	1 ( $f$ )	2 Gaussian simulations

Table 1: Computation effort per timestep

### 3.2 Numerical illustration of strong convergence properties

In order to illustrate the strong convergence rate of a discretization scheme  $\widehat{X}^N$ , we consider the squared  $L^2$ -norm of the supremum of the difference between the scheme with time step  $\frac{T}{N}$  and the one with time step  $\frac{T}{2N}$  :

$$\mathbb{E} \left[ \max_{0 \leq k \leq N} \left| \widehat{X}_{t_k}^N - \widehat{X}_{t_k}^{2N} \right|^2 \right]. \quad (26)$$

This quantity will exhibit the same asymptotic behavior with respect to  $N$  as the squared  $L^2$ -norm of the difference between the scheme with time step  $\frac{T}{N}$  and the limiting process towards which it converges (see Alfonsi [1]).

In Figure 1, we draw the logarithm of the Monte Carlo estimation of (26) as a function of the logarithm of the number of time steps. The number of discretization steps is a power of 2 varying from 2 to 256 and the number of Monte Carlo samples used is equal to  $M = 10\,000$ . We also consider the strong convergence of the schemes on the asset itself (see Figure 2) by computing  $\mathbb{E} \left[ \max_{0 \leq k \leq N} \left| e^{\widehat{X}_{t_k}^N} - e^{\widehat{X}_{t_k}^{2N}} \right|^2 \right]$ .

The confidence intervals of the estimations are reported in error bars in the figures : as one can see, the number of simulations considered suffices to have precise results. The average width of the confidence intervals reported in figures 1 and 2 is equal to 0.07. Note that, since the width of the confidence interval in the estimation of (26) is proportional to the standard error which should theoretically be proportional to  $N$  too, then the width of the confidence interval expressed in logarithmic scale should be constant. We can see that this is indeed the case, especially when the number of time-steps is large enough.

The slopes of the regression lines are reported in Table 2. For completeness sake, we give the standard deviation of the residuals in the regression. We see that, both for the logarithm of the asset and for the asset itself, all the schemes exhibit a strong convergence of order  $\frac{1}{2}$ . Our schemes only have a better constant.

	WeakTraj_1	Weak_2	OU_Improved	IJK	CMT	Euler
Log-asset	-1.01 (0.06)	-0.88 (0.03)	-0.94 (0.04)	-0.92 (0.07)	-0.98 (0.02)	-0.84 (0.08)
Asset	-1.01 (0.06)	-0.91 (0.05)	-0.95 (0.02)	-0.88 (0.08)	-0.95 (0.06)	-0.85 (0.09)

Table 2: Slopes of the regression lines (Strong convergence)

#### 3.2.1 Weak trajectorial convergence

Nevertheless, as explained in Remark 5, for the scheme with time step  $\frac{1}{N}$ , one can replace the increments of the Brownian motion  $(B_t)_{t \in [0, T]}$  by a sequence of Gaussian random variables smartly constructed from the scheme with time step  $\frac{1}{2N}$ . This particular coupling is possible whenever the independence structure between  $(B_t)_{t \in [0, T]}$  and  $(Y_t)_{t \in [0, T]}$  is preserved by the discretization of the latter process, which is the case for all the schemes but the CMT scheme. So we carry out this coupling and we repeat the preceding numerical experiment. The results are put together in Figures 3 and 4 and in Table 3. The average width of the confidence intervals is equal to 0.09.

As expected, we see that the WeakTraj\_1 and the OU\_Improved schemes exhibit a first order convergence rate whereas the other schemes exhibit a  $\frac{1}{2}$  order convergence rate. Note that the CMT scheme has a weak

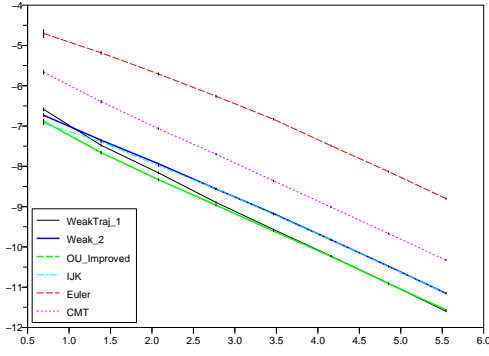


Figure 1: Strong convergence on the log-asset

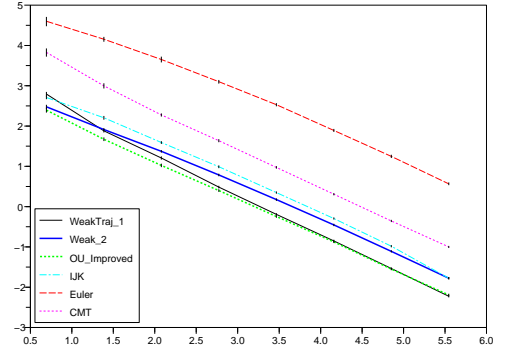


Figure 2: Strong convergence on the asset

trajectorial convergence of order one but it is much more difficult to implement the coupling for which the convergence order is indeed equal to one.

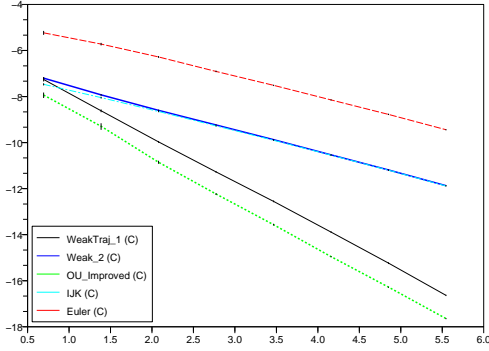


Figure 3: Weak trajectorial convergence on the log-asset (with coupling)

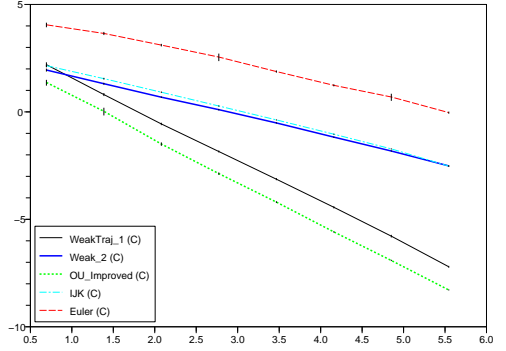


Figure 4: Weak trajectorial convergence on the asset (with coupling)

	WeakTraj_1	Weak_2	OU_Improved	IJK	CMT	Euler
Log-asset	-1.92 (0.03)	-0.91 (0.02)	-1.99 (0.06)	-0.95 (0.03)	—	-0.85 (0.05)
Asset	-1.92 (0.04)	-0.95 (0.03)	-2 (0.05)	-0.91 (0.06)	—	-0.87 (0.09)

Table 3: Slopes of the regression lines (Weak trajectorial convergence)

We repeat the same numerical experiments for the Stein & Stein and the quadratic Gaussian model. The results are reported in figures 5 and 6 and in Table 4. We observe that the theoretical results stated in Proposition (10) are confirmed by the numerical findings : the slope of the regression line is approximately equal to 1.3 for both Weak\_Traj1 and OU\_Improved schemes in the Stein&Stein model whereas for the quadratic Gaussian model, it is approximately equal to 2.

### 3.2.2 Convergence at terminal time

We consider now convergence at terminal time, precisely the squared  $L^2$ -norm of the difference between the terminal values of the schemes with time steps  $\frac{T}{N}$  and  $\frac{T}{2N}$  :

$$\mathbb{E} \left[ \left| \hat{X}_T^N - \hat{X}_T^{2N} \right|^2 \right]. \quad (27)$$

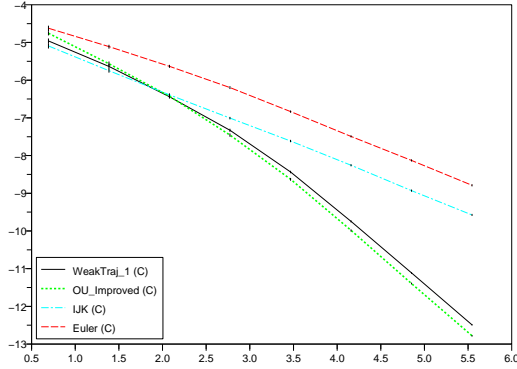


Figure 5: Quadratic Gaussian model - Weak trajectorial convergence on the asset (with coupling)

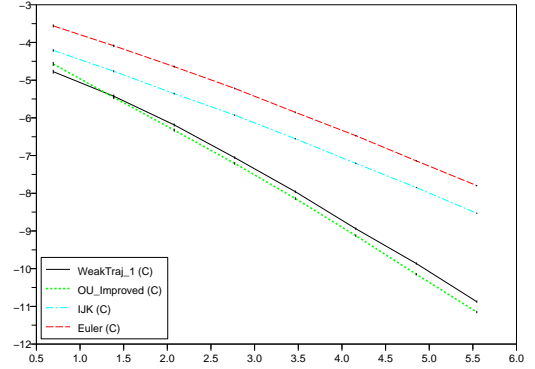


Figure 6: Stein & Stein model - Weak trajectorial convergence on the asset (with coupling)

	WeakTraj_1	OU_Improved	IJK	Euler
Quadratic Gaussian model - Asset	-1.95 (0.02)	-1.99 (0.01)	-0.94 (0.07)	-0.94 (0.01)
Stein&Stein model - Asset	-1.3 (0.12)	-1.35 (0.07)	-0.89 (0.04)	-0.87 (0.06)

Table 4: Slopes of the regression lines (Weak trajectorial convergence) - Quadratic Gaussian and Stein&Stein models

Note that we introduce a coupling : we write the schemes straight at the terminal time as we did for the Weak\_2 scheme (see (24)) and we generate the terminal values of the schemes with time steps  $\frac{T}{N}$  and  $\frac{T}{2N}$  using the same single normal random variable to simulate the stochastic integral w.r.t.  $(B_t)_{t \in [0, T]}$ . Once again, it is possible to proceed alike for all the schemes but the CMT scheme. For the latter, we simulate the scheme at all the intermediate discretization times to obtain the value at terminal time.

We also consider the convergence at terminal time of the asset itself. We report the numerical results in Figures 7 and 8 and give the slopes of the regression lines in Table 5.

	WeakTraj_1	Weak_2	OU_Improved	IJK	CMT	Euler
Log-asset	-2.03 (0.04)	-2 (0.05)	-2.97 (0.03)	-1.97 (0.02)	-1.05 (0.04)	-1.34 (0.19)
Asset	-2.02 (0.04)	-1.98 (0.04)	-2.97 (0.06)	-1.95 (0.03)	-1.08 (0.08)	-1.34 (0.18)

Table 5: Slopes of the regression lines (Convergence at terminal time)

We observe that, as stated in Remark 11, the OU\_Improved scheme exhibits a convergence rate of order  $\frac{3}{2}$ , outperforming all the other schemes. As previously, the WeakTraj\_1 scheme exhibits a first order convergence rate. Note also that this new coupling at terminal time improved the convergence rate of the Weak\_2 and the IJK schemes up to order one and, surprisingly, it improved the convergence rate of the Euler scheme up to an order strictly greater than the expected  $\frac{1}{2}$ , approximately 0.67.

### 3.3 Standard call pricing

#### 3.3.1 Numerical illustration of weak convergence

We compute the price of a call option with strike  $K = 100$  and maturity  $T = 1$ . For all the schemes but the CMT scheme, we use the conditioning variance reduction technique presented in Remark 13.

In Figure 9 we draw the logarithm of the pricing error :  $\log(|P_{\text{exact}} - P_{\text{scheme}}^N|)$  where  $P_{\text{exact}} \approx 12.82603$  is obtained by a multilevel Monte Carlo with an accuracy of  $5 \times 10^{-4}$ , as a function of the logarithm of the number of time steps. In order to avoid statistical noise, we make  $10^7$  simulations.



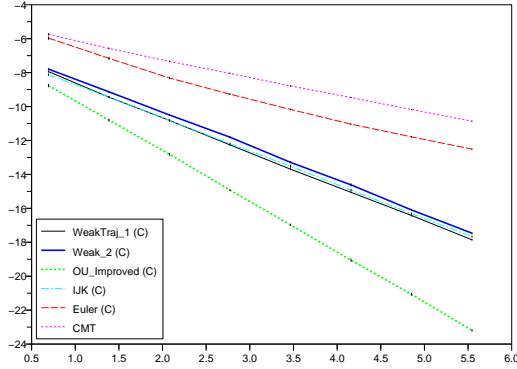


Figure 7: Convergence at terminal time for the log-asset

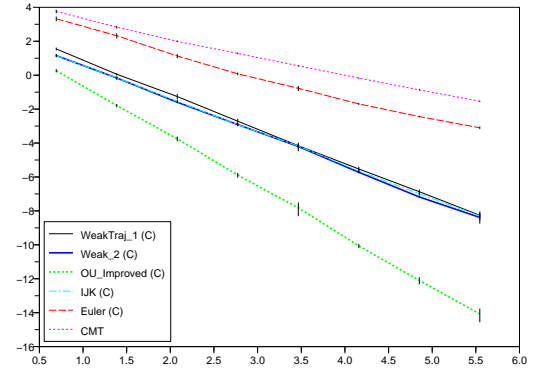


Figure 8: Convergence at terminal time for the asset

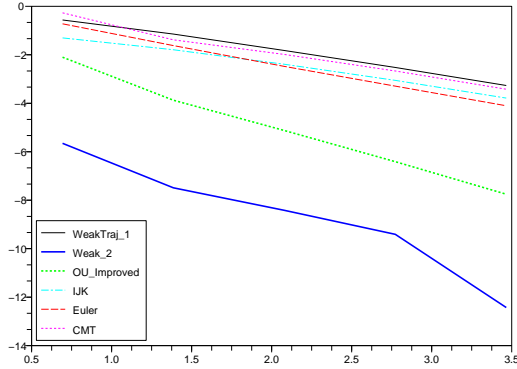


Figure 9: Illustration of the convergence rate for the call option

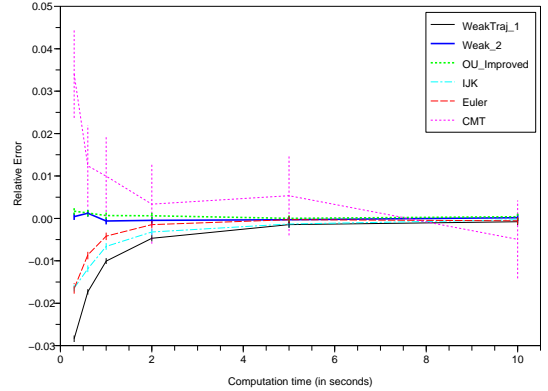


Figure 10: Convergence of the call price with respect to time

We see that, as expected, the Weak\_2 scheme and the OU\_Improved scheme exhibit a weak convergence of order two and converge much faster than the others. The weak scheme already gives an accurate price with only four time steps. The WeakTraj\_1 scheme has a weak convergence of order one like the Euler and the IJK scheme, but it has a greater leading error term. Fortunately, its better strong convergence properties enable it to catch up with the multilevel Monte Carlo method as we will see hereafter.

We also repeat this numerical experiment with the Stein&Stein and the quadratic Gaussian models (see figures 11 and 12) and check that the same conclusions hold.

Finally, note that the weak scheme does not require the simulation of additional terms when compared to the Euler or the IJK schemes. Combined with its second order weak convergence order, this makes the Weak\_2 scheme very competitive for the pricing of plain vanilla European options. In figure 10, we give the relative error of each scheme as a function of the computation time needed when we fix the number of simulations to  $M = 100\,000$ . We see that both the Euler and the IJK scheme take five seconds to reach the relative error obtained with the Weak\_2 scheme and the OU\_Improved in less than a second. Note finally that the confidence interval is much larger for the CMT scheme than for the other schemes because of the use of the conditioning variance reduction technique for these schemes.

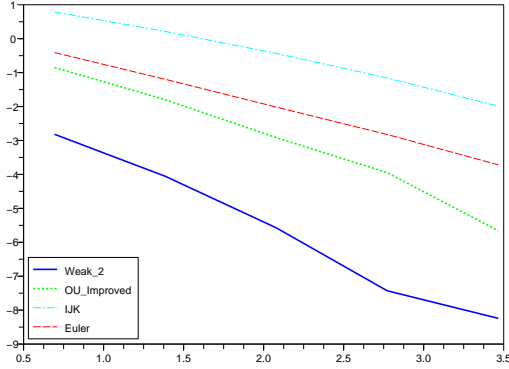


Figure 11: Stein&Stein model - Illustration of the convergence rate for the call option

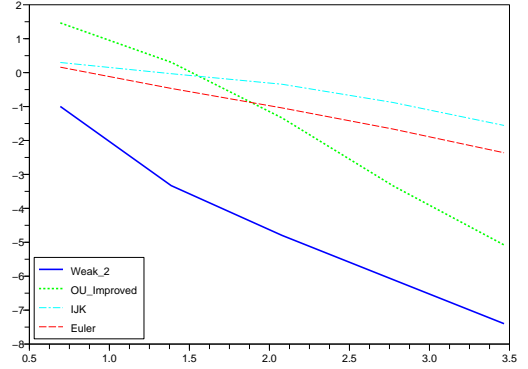


Figure 12: Quadratic Gaussian model - Illustration of the convergence rate for the call option

### 3.3.2 Multilevel Monte Carlo

Let us now apply the multilevel Monte Carlo method of Giles [9] to compute the Call price. As previously, we consider the schemes straight at the terminal time and use a conditioning variance reduction technique. We give the CPU time as a function of the accuracy parameter Epsilon in Figure 13. This accuracy parameter is slightly higher than the root mean square error achieved (see section 4.2 of [9] for details on the heuristic numerical algorithm which is used). We check this numerically by computing different ratios between the root mean square error achieved using the reference value  $P_{\text{exact}}$  and the target accuracy Epsilon (see table 6).

	WeakTraj_1	Weak_2	OU_Improved	IJK	Euler
Epsilon= $10^{-1}$	0.96	0.53	0.6	0.98	0.61
Epsilon= $10^{-2}$	0.8	0.85	0.85	0.94	0.81
Epsilon= $10^{-3}$	0.6	0.57	0.6	0.64	0.7
Epsilon= $10^{-4}$	0.8	0.98	0.54	0.6	0.91

Table 6: Ratio between rmse and Epsilon

Figure 13 shows that both the Weak\_2 and the OU\_Improved scheme are great time-savers. For the OU\_Improved scheme, the effect coming from its good strong convergence properties is somewhat offset by the additional terms that it requires to simulate. We can see nevertheless that it is going to overcome the Weak\_2 scheme for higher accuracy levels.

In order to illustrate the benefits of the multilevel Monte Carlo method, we also give the variation of the computational complexity  $C$ , defined as the total number of timesteps performed on all levels (see section 5 of [9]), with the desired accuracy with and without multilevel for the OU\_Improved scheme (see figure 14).

### 3.4 Lookback option pricing and multilevel Monte Carlo

Finally, we consider an example of path-dependent option pricing : the lookback option. More precisely, we compute the price of the option whose pay-off is equal to  $S_T - \min_{t \in [0, T]} S_t$ . The use of multilevel Monte Carlo for lookback options in local volatility models discretized by the Euler scheme was justified in [11].

In order to take full advantage of the good convergence properties of our schemes, we approximate the minimum of the scheme by the minimum of a drifted Brownian motion. This is similar to what is done in [10].

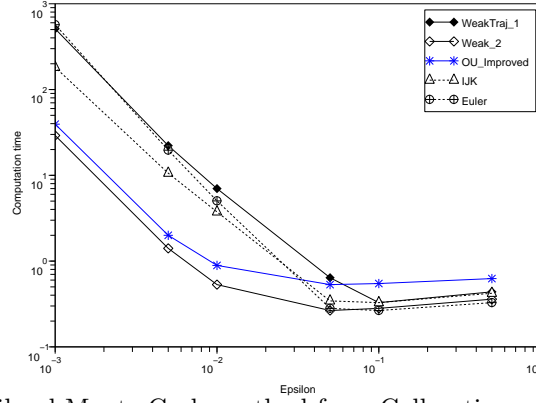


Figure 13: Multilevel Monte Carlo method for a Call option using different schemes

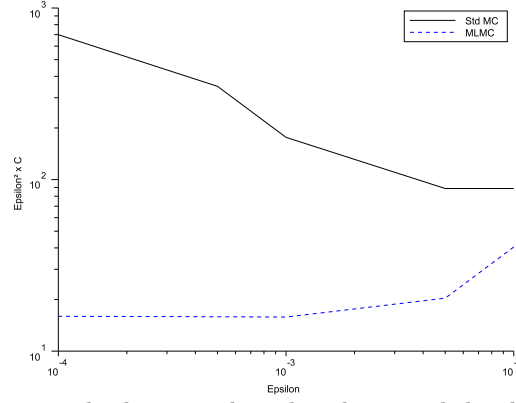


Figure 14: OU\_Improved scheme with and without multilevel Monte Carlo method

More precisely, for the WeakTraj\_1 scheme, consider the interval  $[k\frac{T}{N}, (k+1)\frac{T}{N}]$ .

Scheme with time step  $\delta_{2N}$  :

We approximate  $\min_{t \in [k\frac{T}{N}, (k+1)\frac{T}{N}]} S_t$  by  $\tilde{m}_{2k}^{2N} \wedge \tilde{m}_{2k+1}^{2N}$  where,  $\forall 0 \leq j \leq 2N-1$ ,

$$\tilde{m}_j^{2N} = \frac{1}{2} \left( e^{\tilde{X}_j^{2N} \frac{T}{2N}} + S_{(j+1)\frac{T}{2N}}^{e,2N} - \sqrt{\left( e^{\tilde{X}_j^{2N} \frac{T}{2N}} - S_{(j+1)\frac{T}{2N}}^{e,2N} \right)^2 - 2e^{2\tilde{X}_j^{2N} \frac{T}{2N}} f^2(Y_{j\frac{T}{2N}}) \frac{T}{2N} \ln(U_j)} \right),$$

where  $S_{(j+1)\frac{T}{2N}}^{e,2N} = e^{\tilde{X}_j^{2N} \frac{T}{2N}} \left( 1 + r\frac{T}{2N} + f(Y_{j\frac{T}{2N}}) \left( \rho(W_{(j+1)\frac{T}{2N}} - W_{j\frac{T}{2N}}) + \sqrt{1-\rho^2}(B_{(j+1)\frac{T}{2N}} - B_{j\frac{T}{2N}}) \right) \right)$  and  $(U_j)_{0 \leq j \leq 2N-1}$  is an independent sequence of independent random variable uniformly distributed.

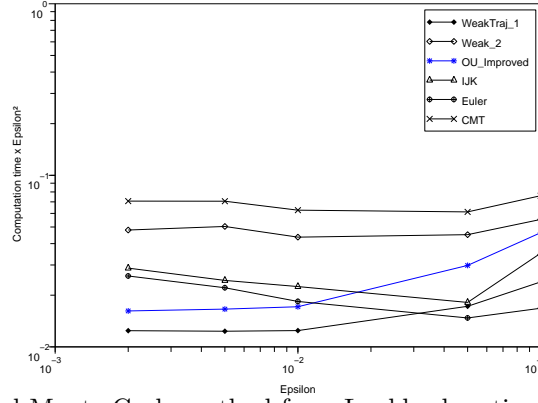


Figure 15: Multilevel Monte Carlo method for a Lookback option using different schemes.

Scheme with time step  $\delta_N$  :

According to Remark 5,  $\tilde{X}_{t_{k+1}}^N$  is computed using the Brownian increment  $\Delta\tilde{B}_{k+1}^N$  given by a linear combination of  $(B_{(2k+1)\frac{T}{2N}} - B_{k\frac{T}{N}}, B_{(k+1)\frac{T}{N}} - B_{(2k+1)\frac{T}{2N}})$  (see (10)). Now, to prevent bias, we are going to approximate  $\min_{t \in [k\frac{T}{N}, (k+1)\frac{T}{N}]} S_t$  by the minimum  $\min_{[t_k, t_{k+1}]} \tilde{S}_t^{e,2N}$  of some Euler scheme  $\tilde{S}_t^{e,2N}$  like in the scheme with time step  $\delta_{2N}$ . To remain consistent, we have to choose

$$\tilde{S}_{t_{k+1}}^{e,2N} = e^{\tilde{X}_{t_k}^N} \left( 1 + r\frac{T}{N} + f(Y_{k\frac{T}{N}}) \left( \rho(W_{t_{k+1}} - W_{t_k}) + \sqrt{1 - \rho^2} \Delta\tilde{B}_{k+1}^N \right) \right).$$

In order to ensure a good strong coupling with the scheme with time step  $\delta_{2N}$ , we need to compute the intermediate value  $\tilde{S}_{(2k+1)\frac{T}{2N}}^{e,2N} = e^{\tilde{X}_{t_k}^N} \left( 1 + r\frac{T}{N} + f(Y_{k\frac{T}{N}}) \left( \rho(W_{t_{k+1}} - W_{t_k}) + \sqrt{1 - \rho^2} \Delta\tilde{B}_{2k+1}^{2N} \right) \right)$  using some Brownian increment  $\Delta\tilde{B}_{2k+1}^{2N}$  as close as possible to  $B_{(2k+1)\frac{T}{2N}} - B_{k\frac{T}{N}}$  but such that  $\Delta\tilde{B}_{k+1}^N - \Delta\tilde{B}_{2k+1}^{2N}$  is independent of  $\tilde{B}_{2k+1}^{2N}$  and distributed according to  $\mathcal{N}(0, \frac{T}{2N})$ . Choosing  $\Delta\tilde{B}_{2k+1}^{2N}$  of the form  $\frac{a(B_{(2k+1)\frac{T}{2N}} - B_{k\frac{T}{N}}) + b(B_{(k+1)\frac{T}{N}} - B_{(2k+1)\frac{T}{2N}})}{\sqrt{a^2 + b^2}}$  and maximizing  $Cov\left(\Delta\tilde{B}_{2k+1}^{2N}, B_{(2k+1)\frac{T}{2N}} - B_{k\frac{T}{N}}\right) = \frac{a}{\sqrt{a^2 + b^2}}$  leads to  $a = v_{2k}^{2N} + v_{2k+1}^{2N}$  and  $b = v_{2k+1}^{2N} - v_{2k}^{2N}$  (see Remark 5 for the definition of  $v^{2N}$ ).

Finally, we approximate  $\min_{t \in [k\frac{T}{N}, (k+1)\frac{T}{N}]} S_t$  by  $\tilde{m}_k^N \wedge \tilde{m}_{k+1}^N$  where

$$\tilde{m}_k^N = \frac{1}{2} \left( e^{\tilde{X}_{k\frac{T}{N}}^N} + \tilde{S}_{(2k+1)\frac{T}{2N}}^{e,2N} - \sqrt{\left( e^{\tilde{X}_{k\frac{T}{N}}^N} - \tilde{S}_{(2k+1)\frac{T}{2N}}^{e,2N} \right)^2 - 2e^{2\tilde{X}_{k\frac{T}{N}}^N} f^2(Y_{k\frac{T}{N}}) \frac{T}{2N} \ln(U_{2k})} \right)$$

and

$$\tilde{m}_{k+1}^N = \frac{1}{2} \left( \tilde{S}_{(2k+1)\frac{T}{2N}}^{e,2N} + \tilde{S}_{(k+1)\frac{T}{N}}^{e,2N} - \sqrt{\left( \tilde{S}_{(2k+1)\frac{T}{2N}}^{e,2N} - \tilde{S}_{(k+1)\frac{T}{N}}^{e,2N} \right)^2 - 2e^{2\tilde{X}_{k\frac{T}{N}}^N} f^2(Y_{k\frac{T}{N}}) \frac{T}{2N} \ln(U_{2k+1})} \right).$$

The numerical results we obtain are very satisfactory. In figure 15, we draw the CPU time multiplied by the mean square error against the root mean square error. We see that our schemes perform much better than the others.

## 4 Conclusion

In this article, we have capitalized on the particular structure of stochastic volatility models to propose and discuss two simple and yet competitive discretization schemes. The first one exhibits first order weak trajectorial convergence and has the advantage of improving multilevel Monte Carlo methods for the pricing of path dependent options. The second one is rather useful for pricing European options since it has a second order weak convergence rate.

We have also focused on the special case of an Ornstein-Uhlenbeck process driving the volatility, which encompasses many stochastic volatility models such as the Scott's model [27] or the quadratic Gaussian model. Then, the convergence properties of the previous schemes are preserved when simulating  $(Y_t)_{0 \leq t \leq T}$  exactly. We have also proposed an improved scheme exhibiting both weak trajectorial convergence of order one and weak convergence of order two.

Our numerical experiments confirm the theoretical rates of convergence of our schemes. We also compare the time needed by the different schemes to achieve a given precision in the multilevel Monte Carlo computation of a plain vanilla Call option and a lookback option. For high levels of precision our schemes turn out to be more efficient than the Euler, the Kahl-Jäckel and the Cruzeiro-Malliavin-Thalmaier schemes for both the vanilla Call and the lookback option. The reason is that their better convergence properties compensate the increase of computation effort at each step.

As a last remark, we point out that our results can be naturally extended to stochastic volatility models where the constant correlation coefficient is replaced by a function  $\rho(Y_t)$  of the process driving the stochastic volatility in (1). In this case, if one considers the transformation  $F(y) = \int^y \frac{\rho(z)f(z)}{\sigma(z)} dz$  and carries out the same analysis then one should obtain weak trajectorial convergence results under additional regularity assumptions on the function  $\rho$ .

## References

- [1] A. Alfonsi. On the discretization schemes for the CIR (and Bessel squared) processes. *Monte Carlo Methods and Applications*, 11(4):355–384, 2005.
- [2] A. Alfonsi. High order discretization schemes for the CIR process: Application to affine term structure and Heston models. *Mathematics of Computations*, 79:209–237, 2009.
- [3] L. Andersen. Efficient simulation of the heston stochastic volatility model. *SSRN eLibrary*, 2007.
- [4] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probability Theory and Related Fields*, 104(1):43–60, 1996.
- [5] A. Berkaoui, M. Bossy, and A. Diop. Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence. *ESAIM. Probability and Statistics*, 12:1–11, 2008.
- [6] M. Broadie and Ö Kaya. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, 54(2):217–231, 2006.
- [7] A.B. Cruzeiro, P. Malliavin, and A. Thalmaier. Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, 338(6):481–486, 2004.
- [8] G. Deelstra and F. Delbaen. Convergence of discretized stochastic (interest rate) processes with stochastic drift term. *Applied Stochastic Models and Data Analysis*, 14(1):77–84, 1998.
- [9] M. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
- [10] M. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and quasi-Monte Carlo methods 2006*, pages 343–358. Springer, Berlin, 2008.

- [11] M. Giles, D. Higham and X. Mao. Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff. *Finance and Stochastics*, 13(3):403–413, 2009.
- [12] J. Guyon. Euler scheme and tempered distributions. *Stochastic Processes and their Applications*, 116(6):877–904, 2006.
- [13] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300, 1987.
- [14] C. Kahl and P. Jäckel. Fast strong approximation Monte Carlo schemes for stochastic volatility models. *Quantitative Finance*, 6(6):513–536, 2006.
- [15] C. Kahl and H. Schurz. Balanced Milstein methods for ordinary SDEs. *Monte Carlo Methods and Applications*, 12(2):143–170, 2006.
- [16] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus*. Springer-Verlag New-York, 2nd edition, 1991.
- [17] A. Kebaier. Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. *The Annals of Applied Probability*, 15(4):2681–2705, 2005.
- [18] S. Kusuoka. Approximation of expectation of diffusion process and mathematical finance. *Taniguchi Conference on Mathematics Nara '98*, 31:147–165, 2001.
- [19] S. Kusuoka. Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus. *Advances in mathematical economics*, 6:69–83, 2004.
- [20] B. Lapeyre and E. Temam. Competitive Monte Carlo methods for pricing Asian options. *Journal of Computational Finance*, 5(1), 2001.
- [21] R. Lord, R. Koekkoek, and D.J. Van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *SSRN eLibrary*, 2008.
- [22] T. Lyons and N. Victoir. Cubature on Wiener space. *Proceedings of The Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences*, 460(2041):169–198, 2004.
- [23] G.N. Milstein. *Numerical Integration of Stochastic Differential Equations*, volume 313. Kluwer Academic Publishers, 1995.
- [24] S. Ninomiya and M. Ninomiya. A new higher-order weak approximation scheme for stochastic differential equations and the Runge-Kutta method. *Finance and Stochastics*, 13:415–443, 2009.
- [25] S. Ninomiya and N. Victoir. Weak approximation of stochastic differential equations and application to derivative pricing. *Applied Mathematical Finance*, 15(1-2):107–121, 2008.
- [26] M. Romano and N. Touzi. Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance*, 7(4):399–412, 1997.
- [27] L.O. Scott. Option pricing when the variance changes randomly: theory, estimation, and an application. *The Journal of Financial and Quantitative Analysis*, 22(4):419–438, 1987.
- [28] E.M. Stein and J.C. Stein. Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies*, 4(4):727–752, 1991.
- [29] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8(4):483–509, 1990.

- [30] H. Tanaka and A. Kohatsu-Higa. An operator approach for Markov chain weak approximations with an application to infinite activity Lévy driven SDEs. *The Annals of Applied Probability*, 19(3):1026–1062, 2009.
- [31] J.B. Wiggins. Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics*, 19(2):351–372, 1987.

## 5 Appendix

### 5.1 Proof of Lemma 3

We first suppose that  $p = 1$ . According to Theorem 5.2 page 72 of Milstein [23], it suffices to check that there exists a positive constant  $C$  independent of  $N$  such that

$$\begin{aligned} \left| \mathbb{E} \left( Y_{\delta_N} - \bar{Y}_{\delta_N}^N \right) \right| &\leq C \delta_N^2 \\ \left| \mathbb{E} \left( \left( Y_{\delta_N} - \bar{Y}_{\delta_N}^N \right)^2 \right) \right|^{\frac{1}{2}} &\leq C \delta_N^{\frac{3}{2}} \\ \left| \mathbb{E} \left( \left( Y_{\delta_N} - \bar{Y}_{\delta_N}^N \right)^4 \right) \right|^{\frac{1}{4}} &\leq C \delta_N^{\frac{5}{4}}. \end{aligned} \tag{28}$$

First note that

$$Y_{\delta_N} - \bar{Y}_{\delta_N}^N = \int_0^{\delta_N} b(Y_s) - b(y_0) ds + \int_0^{\delta_N} \left( \int_0^s (b\sigma' + \frac{1}{2}\sigma^2\sigma'')(Y_r) dr + (\sigma\sigma'(Y_r) - \sigma\sigma'(y_0)) dW_r \right) dW_s.$$

Thanks to Itô's formula and to assumption  $(\mathcal{H}5)$ , we have that

$$\begin{aligned} \left| \mathbb{E} \left( Y_{\delta_N} - \bar{Y}_{\delta_N}^N \right) \right| &= \left| \int_0^{\delta_N} \int_0^s \mathbb{E} \left( (bb' + \frac{1}{2}b''\sigma^2)(Y_r) \right) dr ds \right| \\ &\leq C \left| \int_0^{\delta_N} \int_0^s C(1 + \mathbb{E}(|Y_r|^2)) dr ds \right| \\ &\leq C \delta_N^2. \end{aligned}$$

Using assumptions  $(\mathcal{H}5)$  and  $(\mathcal{H}6)$ , we also have  $\forall p \geq 1$

$$\begin{aligned} \mathbb{E} \left( \left| Y_{\delta_N} - \bar{Y}_{\delta_N}^N \right|^{2p} \right) &\leq 2^{2p-1} \mathbb{E} \left[ \left| \int_0^{\delta_N} b(Y_s) - b(y_0) ds \right|^{2p} \right. \\ &\quad \left. + \left| \int_0^{\delta_N} \left( \int_0^s (b\sigma' + \frac{1}{2}\sigma\sigma'')(Y_r) dr + (\sigma\sigma'(Y_r) - \sigma\sigma'(y_0)) dW_r \right) dW_s \right|^{2p} \right] \\ &\leq 2^{2p-1} \left[ \delta_N^{2p-1} \int_0^{\delta_N} \mathbb{E} (|b(Y_s) - b(y_0)|^{2p}) ds \right. \\ &\quad \left. + C \delta_N^{p-1} \int_0^{\delta_N} \mathbb{E} \left( \left| \int_0^s (b\sigma' + \frac{1}{2}\sigma\sigma'')(Y_r) dr + (\sigma\sigma'(Y_r) - \sigma\sigma'(y_0)) dW_r \right|^{2p} \right) ds \right] \\ &\leq C \left[ \delta_N^{2p-1} \int_0^{\delta_N} s^p ds + \delta_N^{p-1} \int_0^{\delta_N} s^{2p-1} \int_0^s \mathbb{E} \left( \left| (b\sigma' + \frac{1}{2}\sigma\sigma'')(Y_r) \right|^{2p} \right) dr ds \right. \\ &\quad \left. + \delta_N^{p-1} \int_0^{\delta_N} s^{p-1} \int_0^s \mathbb{E} (|\sigma\sigma'(Y_r) - \sigma\sigma'(y_0)|^{2p}) dr ds \right] \\ &\leq C \delta_N^{3p}. \end{aligned}$$

This implies both the second and the third inequality of (28). This estimation is also sufficient to extend the result of Milstein [23] to the  $L^{2p}$  norm and conclude the proof.

## 5.2 Proof of Lemma 8

One can easily check that  $(Y_t)_{0 \leq t \leq T}$  is a Gaussian process which has the same distribution law as the process  $(y_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t}) + \frac{\nu e^{-\kappa t}}{\sqrt{2\kappa}} W_{e^{2\kappa t} - 1})_{0 \leq t \leq T}$ . So,

$$\begin{aligned} \mathbb{E} \left( e^{c_1 \sup_{0 \leq t \leq T} |Y_t|^{1+c_2}} \right) &= \mathbb{E} \left( e^{c_1 \sup_{0 \leq t \leq T} |y_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t}) + \frac{\nu e^{-\kappa t}}{\sqrt{2\kappa}} W_{e^{2\kappa t} - 1}|^{1+c_2}} \right) \\ &\leq C \mathbb{E} \left( e^{C \sup_{0 \leq t \leq T} |W_{e^{2\kappa t} - 1}|^{1+c_2}} \right). \end{aligned}$$

Since  $\sup_{0 \leq t \leq e^{2\kappa T} - 1} |W_t| = (\sup_{0 \leq t \leq e^{2\kappa T} - 1} W_t) \vee (-\inf_{0 \leq t \leq e^{2\kappa T} - 1} W_t)$ , we deduce from the symmetry property of the Brownian motion that

$$\begin{aligned} \mathbb{E} \left( e^{c_1 \sup_{0 \leq t \leq T} |Y_t|^{1+c_2}} \right) &\leq C \mathbb{E} \left( e^{C |\sup_{0 \leq t \leq e^{2\kappa T} - 1} W_t|^{1+c_2}} + e^{C |\inf_{0 \leq t \leq e^{2\kappa T} - 1} W_t|^{1+c_2}} \right) \\ &\leq 2C \mathbb{E} \left( e^{C |\sup_{0 \leq t \leq e^{2\kappa T} - 1} W_t|^{1+c_2}} \right). \end{aligned}$$

The probability density function of  $\sup_{0 \leq t \leq T} W_t$  is equal to  $y \mapsto \sqrt{\frac{2}{\pi T}} e^{-\frac{y^2}{2T}} \mathbb{1}_{\{y > 0\}}$  (see for example problem 8.2 p. 96 of Karatzas and Shreve [16]) which permits to conclude.

Let us now assume that  $y_0 \neq 0$ . Then  $t_0 \stackrel{\text{def}}{=} \inf\{t \in [0, T] : y_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t}) = 0\}$  (convention  $\inf \emptyset = T$ ) is positive.  $\mathbb{P}[|Y_0| \leq N^{-\alpha}] = 0$  for  $N$  large enough and for  $t \in (0, T]$ ,

$$\mathbb{P}[|Y_t| \leq N^{-\alpha}] \leq \frac{2N^{-\alpha}}{\nu} \sqrt{\frac{\kappa}{\pi(1 - e^{-2\kappa t})}} \exp \left( -\frac{\kappa(|y_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t})| - N^{-\alpha})^2}{\nu^2(1 - e^{-2\kappa t})} \right).$$

Since  $\lim_{N \rightarrow \infty} \inf_{t \in [0, \frac{t_0}{2}]} (|y_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t})| - N^{-\alpha}) > 0$ , one deduces that  $\sup_{t \in [0, \frac{t_0}{2}]} \mathbb{P}[|Y_t| \leq N^{-\alpha}] \leq CN^{-\alpha}$ . The same conclusion holds for  $\sup_{t \in [\frac{t_0}{2}, T]} \mathbb{P}[|Y_t| \leq N^{-\alpha}]$  by bounding the exponential factor by 1.